applied sciences

MDPI

Article

A Real-Time Remote Respiration Measurement Method with Improved Robustness Based on a CNN Model

Hyeonsang Hwang [1,†], Kunyoung Lee [1,†] and Eui Chul Lee [2,*]

[1] Department of Computer Science, Graduate School, Sangmyung University, Hongjimun 2-Gil 20, Jongno-Gu, Seoul 03016, Republic of Korea
[2] Department of Human-Centered Artificial Intelligence, Sangmyung University, Hongjimun 2-Gil 20, Jongno-Gu, Seoul 03016, Republic of Korea
* Correspondence: eclee@smu.ac.kr; Tel.: +82-2-781-7553
† These authors contributed equally to this work.

Abstract: Human respiration reflects meaningful information, such as one's health and psychological state. Rates of respiration are an important indicator in medicine because they are directly related to life, death, and the onset of a serious disease. In this study, we propose a noncontact method to measure respiration. Our proposed approach uses a standard RGB camera and does not require any special equipment. Measurement is performed automatically by detecting body landmarks to identify regions of interest (RoIs). We adopt a learning model trained to measure motion and respiration by analyzing movement from RoI images for high robustness to background noise. We collected a remote respiration measurement dataset to train the proposed method and compared its measurement performance with that of representative existing methods. Experimentally, the proposed method showed a performance similar to that of existing methods in a stable environment with restricted motion. However, its performance was significantly improved compared to existing methods owing to its robustness to motion noise. In an environment with partial occlusion and small body movement, the error of the existing methods was 4–8 bpm, whereas the error of our proposed method was around 0.1 bpm. In addition, by measuring the time required to perform each step of the respiration measurement process, we confirmed that the proposed method can be implemented in real time at over 30 FPS using only a standard CPU. Since the proposed approach shows state-of-the-art accuracy with the error of 0.1 bpm in the wild, it can be expanded to various applications, such as medicine, home healthcare, emotional marketing, forensic investigation, and fitness in future research.

Keywords: noncontact monitoring; remote sensing; vital signs; respiratory rate estimation

Citation: Hwang, H.; Lee, K.; Lee, E.C. A Real-Time Remote Respiration Measurement Method with Improved Robustness Based on a CNN Model. *Appl. Sci.* **2022**, *12*, 11603. https://doi.org/10.3390/app122211603

Academic Editors: Hyeonjoon Moon and Irfan Mehmood

Received: 27 October 2022
Accepted: 12 November 2022
Published: 15 November 2022

1. Introduction

Respiration is a basic vital sign that can be used to intuitively evaluate human health along with heart rate, blood pressure, and body temperature. It is widely used as a standard indicator in medical, human–computer interaction, forensic investigation, and sports research [1–6]. In addition, changes in respiration are important to understand changes in the vital state of the body because levels of oxygen and carbon dioxide in the blood suitable for life activities are maintained by controlling the rate of respiration [7]. Respiration can be used as a good indicator to monitor vital signs, and monitoring respiration rates can also help identify high-risk patients and prevent dangerous situations because physical changes due to respiration exhibit greater variability compared to those caused by heart rate and blood pressure [8,9]. Even in medical environments, increased rates of respiration are effective in predicting various abnormal conditions, such as respiratory failure due to muscle weakness, readmission to intensive care units, and cardiopulmonary arrest, and 21% of patients with 25–29 breaths per minute die in the hospital [10,11]. Therefore, research and development on vital sign monitoring systems that include respiratory rate measurement can be usefully utilized to prevent physical health risks by detecting deterioration in patients' heart conditions in advance [12,13]. Respiration is influenced not only by physical health, but also by psychological conditions, including anxiety, depression, anger, and stress, and is closely related to changes in levels of consciousness caused by drugs and sleep [14–16]. Therefore, respiration is a very important indicator that can be applied to observe changes in human physical and psychological health and other conditions, and can be widely used in medicine, as well as various other fields. During human metabolism, all cells in the body use the oxygen in the blood to metabolize energy and release carbon dioxide through cellular respiration. As a result of cellular respiration, blood oxygen concentration decreases and carbon dioxide concentration increases, and ventilation occurs; that is, gas is exchanged in the alveoli of the lungs owing to the difference in partial pressure of the gas with air introduced from the outside. To maintain an appropriate level of oxygen concentration, the human body reduces the pressure in the chest cavity by contracting the diaphragm to bring in external air (inhalation) or increases the pressure in the chest cavity by relaxing the diaphragm to expel internal air from the body (exhalation). This process is known as respiration. Thus, respiration can be defined as a complex interaction between the central nervous system, respiratory-related motor neurons, and respiratory muscles [17].

Contact-based methods to measure respiration can be classified into four approaches, including manual measurements [18,19], measuring changes in impedance using electrodes [20–23], measuring pressure using belt-type sensors [24–26], and measuring airflow from the nose and mouth [27–29]. First, a conventional method of manually counting the rate of respiration by eye is widely used. Table 1 summarizes the advantages and disadvantages of each method. These approaches share several limitations in common. Separate sets of special equipment are necessary for each, which are not usually available. In addition, for accurate measurement, the equipment must be in close contact with the patient; the process of preparing to perform the measurement is long and complex, and tends to be inconvenient for the subject, such as by imposing strong pressure or limiting their freedom of movement. Discomfort felt by users may cause inaccurate measurement results owing to changes in on respiration rate occurring for that reason. Although this limitation can be used to understand various human conditions, it limits the application of measuring respiration in various fields. However, changes in the human body due to respiration may not necessarily be observable with contact methods. The movement of the human body due to respiration can be detected by observing the amount of change in the pixels of a video based on optical analysis of motion using motion detection technology, such as optical flow and object detection, performed by applying technology such as frequency analysis based on radar. Changes in temperature due to respiration can also be measured and can be observed by quantitatively measuring changes in temperature in images taken using a thermal imaging camera. In addition, respiration can also be measured based on optical pulse waves through the change in blood flow caused by the change in chest pressure due to respiration [30]. These noncontact methods can overcome some limitations on patients' freedom and enable respiration to be measured more conveniently than with contact methods. Thermal imaging cameras and radar-based methods also involve significant limitations owing to the necessity of using expensive and bulky equipment. However, vision-based technologies using standard, commonly available RGB cameras, such as smartphones and laptops, can also overcome these limitations. Therefore, in this study, we limit the scope of our investigation to remote measurements of respiration using RGB cameras only. Compared with contact respiration measurement methods, noncontact measurements involve several important advantages.

First, owing to the noncontact measurement environment, the existing problems of limited freedom of movement or discomfort can be solved. This provides a more relaxed measurement environment and prevents the occurrence of unnatural breathing patterns by minimizing patients' discomfort.

Table 1. Pros and cons of existing respiration measurement approaches.

Approach	Pros	Cons
Manual measurement [18,19]	No device required	High possibility of human error when monitoring for long periods
Impedance change measurement [20–23]	Various physiological signals can be measured Long-term monitoring possible	Sensitive to posture and movement Special device required
Pressure change measurement [24–26]	Allows for more freedom of movement Relatively strong in movement Long-term monitoring possible	Discomfort caused by physical pressure Special device required
Airflow measurement [27–29]	Accurate measurement Long-term monitoring possible	Restricted environment Special device required

Second, there is no need for specialized equipment to measure respiration only. Existing devices for contact respiration measurement are designed for respiration measurement; therefore, they are not used routinely and are difficult to utilize for other purposes. In contrast, the RGB camera-based respiration measurement method uses technology that is commonly encountered daily and can be used for various purposes; this can improve accessibility because of the specificity of the device.

Third, a cumbersome preparation process for measurement is not required. For the contact respiration measurement methods, a process must be performed to prepare to attach the sensor to the body for measurement, and some basic knowledge is required to attach the sensor correctly. Because the camera-based noncontact respiration measurement device does not require a separate sensor attachment, the preparation process for the measurement can be greatly simplified.

Finally, it can be used as an alternative to contact sensors for patients with skin diseases, children, people with difficulty communicating, and patients with immune disorders who may encounter difficulty attaching sensors to their body. Existing noncontact respiration measurement methods perform image capture in highly controlled situations to achieve sufficient performance, which enables stable measurement of respiration without major impacts on region of interest (RoI) detection performance or signal quality degradation due to noise in the detected RoIs. However, these methods have the disadvantages of being semi-automated and require an environment in which patients' motion is restricted. It is natural for various noise elements to occur in an uncontrolled environment, and thus these methods should be adapted to be used without any additional restrictions to allow stable respiration measurement from the images of the environment that contain noise elements. Therefore, in this study, we propose an RGB camera-based noncontact respiration measurement method that satisfies the following conditions to mitigate the limitations of existing methods and to apply noncontact respiration measurement universally.

1. Noncontact methods can be used to measure respiration without any attached device.
2. We apply stable RoI detection technology to automate the entire respiration measurement process.
3. A breathing signal that is strong against noise elements, such as movement and occlusion noise, can be obtained. The proposed method shows a state-of-the-art accuracy of 0.1 bpm even in the noise environment, which is a noticeable improvement, considering that the existing methods have an error of 4~8 bpm in the noise environment. These improvements can be confirmed through the experimental results in Section 4.
4. The proposed method can operate in real time to immediately detect changes in patients' states. In addition, even in a CPU environment, if there is an RGB camera, it has a measurement speed of 30 FPS or more and is a robust measurement method against motion and occlusion noise, so it can be applied to various fields, such as medicine, home healthcare, emotional marketing, forensic investigation, and fitness.

2. Related Works

2.1. Photoplethysmography-Based Respiration Measurement Method

Both respiration and pulse signals can be observed or extracted from photoplethysmography signals. The decrease in intrathoracic pressure that occurs during inspiration affects venous blood pressure and ventricular volume, which promotes venous blood return and changes cardiac output [31]. In addition, increased intrathoracic pressure due to respiration stimulates the vagus nerve and the sinus node that controls the heart rate, causing respiratory sinus arrhythmia, a phenomenon that lowers the heart rate [32]. Photoplethysmography (PPG) is based on the optical property that the circulation changes the amount of hemoglobin molecules and proteins in the blood, thereby changing the degree of light absorption and scattering across the light spectrum. Therefore, PPG technology can be used to measure the degree of changes in capillary blood due to heartbeat through an optical sensor attached to the surface of the skin [33]. Even without using a sensor attached to the skin surface, these pulse changes can be measured in a noncontact approach using an optical sensor, such as an RGB webcam [34,35]. The acquired remote PPG signal includes amplitude and period changes due to respiration, and respiration can be measured using a noncontact method by analyzing changes in heart rate components due to respiration. Cases and previous studies verified that information on respiration can be estimated using this method [36,37]. In 2010, Poh et al. separated PPG signal and noise components using independent component analysis (ICA) of the average change in brightness of the signal of each RGB channel in the face region and extracted respiration from the power spectrum of the heart rate variability, which is the change in the interval between heartbeats [38–40]. However, although the method of estimating respiration rate using heart rate variability may be effective for young and healthy subjects, it is difficult to apply to older patients or those with chronic disease [41]. In 2011, Yu et al. demonstrated measuring respiration rate in a sports situation by applying a change in the reflected light component of the forehead and blind source separation (BSS) for continuous monitoring of respiration [42]. In addition, Wei et al. obtained the average of each component of the face and each RGB component of the neck region and evaluated the performance of a BSS-based respiration measurement method of the signal of the changes in the average of eight subjects [43]. Tarassenko et al. detected a face region from video and classified the face, upper body, and background using a nonparametric Bayesian image segmentation algorithm and then used an autoregressive model to recognize RGB changes in the face region to estimate respiration information from the corresponding signal [44]. In 2019, Ghodratigohar applied independent component analysis (ICA) to each RGB component of face images to separate the noise and PPG components and decomposed the signal into intrinsic mode functions through complete ensemble empirical mode decomposition with adaptive noise, as well as considerably improved the performance of respiration measurement by selecting the signal that best represented the true respiration [45]. In addition, Sanyal et al. reviewed a method to convert a detected face image into an HSV color model to measure respiration with robustness of changes in light and measured respiration using changes in the color component of a color model. They conducted an experiment on 25 adult subjects (15 men and 10 women) and showed that their approach may be considered promising for application in telemedicine [46]. In 2016, Gastel et al. tracked feature points in an image of the face to improve measurement performance in visible light, as well as infrared environments, by observing changes in the components of light reflected by the skin and estimating signals through the weighted sum of the changes. They proposed a method to measure respiration by applying frequency analysis technology and demonstrated notable performance in visible and infrared environments [47]. PPG-based noncontact respiration measurement methods were also studied. These promising methods can stabilize skin area tracking and continuously measure respiration through a combination of techniques for detecting the facial RoI as a representative skin area. Previous studies showed the usefulness of PPG-based respiration measurement techniques using camera systems. However, these methods still involve some limitations. First, they are difficult to apply in actual environments because most of the research results reported were recorded in environments such as laboratories and hospitals, where both movement and illumination are extremely limited. In fact, techniques such as BSS and frequency analysis methods used to purify respiration information from signals are extremely sensitive to noise; therefore, they cannot estimate respiration rate accurately in the presence of movement and optical noise unrelated to respiration [48]. In addition, Nam et al. and Karlen et al. found that respiratory and heart rate signals were similar to PPG signals when respiratory rates were measured at high respiration rates greater than 20 bpm (bpm). They found that the accuracy was reduced because respiratory rate extraction was impossible owing to the overlapping periodicity [49,50]. Considering that these studies were conducted based on relatively stable PPG signals acquired with a finger in contact with a camera lens and that breathing rates as high as 40 bpm are generally considered normal, the probability of noise is very high. Thus, noncontact PPG-based respiration measurement methods still involve some notable limitations on their application as a general respiration measurement technology [47].

2.2. Motion-Based Respiration Measurement Method

Respiration can also be measured using a noncontact method in which an RGB camera is used to quantitatively detect minute movements of the human body, such as expansion and contraction of the chest caused by respiration. As shown in Figure 1, changes such as the diaphragm moving up and down during breathing, the external intercostal muscles pulling up the ribs and expanding, and the volume of the chest cavity increasing, cause the upper body to move overall [51]. Studies were conducted to measure respiration in a noncontact manner by quantitatively measuring such changes in chest movement. Zhao et al. detected the approximate position of the upper body using the Haar-like feature-based object detection method [52] proposed by Lienhart and Maydt in 2013, and defined body regions as RoIs based on the position of the face to estimate respiration rate by applying frequency analysis techniques to the average change in brightness of a time-series signal within the RoI [53]. Similarly, in 2016, Reyes et al. captured images of the abdomen of the human body using a camera on an Android smartphone and estimated respiration information using the average changes in the pixels of the images [54,55]. Prathosh et al. proposed a method to measure respiration by tracking changes in reflected light using the principle that changes caused by respiratory movement change the light reflected from the chest wall [56]. Massaroni et al. conducted a study to compare the respiration information measured by the average pixel change obtained from RoIs with a reference by selecting the boundary region of the RoI [57]. Although these studies showed that movement caused by respiration can be optically observed, several challenges remain to be solved before these methods can be universally adopted. Estimating motion information by simply using the amount of pixel change is simple to implement, and the computational cost is very low. Figure 2 shows examples of different pixel variation signals that can be observed with the same respiration periods. The scale of the pixel value also varies greatly depending on color changes, and Figure 2 shows that the normalized signal clearly confirms these trends. As may be observed from Figure 2, the pixel variation signal may have a different phase from the actual respiration signal owing to numerous factors, such as the subject's clothes, background, and lighting. In addition, this change may exhibit a completely different periodicity from breathing movement in some cases. These factors can significantly degrade the performance of respiration motion estimation methods based on the amount of change in pixels, and act as a limit that renders detecting the exact timing of inspiration and expiration through motion information impossible. Al-Naji et al. and Antognoli et al. used Eulerian video magnification [58], a technique for amplifying small movements with periodicity, to amplify movement by respiration and then to amplify the difference between adjacent frames [59,60]. However, these methods also perform amplification based on the amount of pixel change; therefore, they are subject to the limitations of pixel change-based methods. To quantitatively measure movement driven by respiration, an optical flow method was used to calculate the movement speed of a specific pixel pattern in an image [61]. Because optical flow estimates a motion vector owing to movement, it can mitigate some problems of methods that simply measure the amount of change in the pixels. In 2016, Lin et al. estimated the relative position of body regions through face detection and measured respiration by calculating the movement of the body regions as an optical flow [62]. In 2018, Tran et al. estimated body RoIs using a depth camera, mapped the image acquired with an RGB camera, and then applied optical flow to the RoI to measure respiration [63]. Janssen et al. obtained a motion matrix by applying optical flow to a video taken in 2015 and considered the respiration characteristics needed to detect RoIs related to respiration in units of pixels by assigning a score, and then estimated respiration information by fusing motion information [64]. They also showed that the distance from the camera or the generation of noise caused by the lighting environment can reduce measurement performance. In addition, Massaroni et al. in 2019 confirmed that the optical flow-based method showed better performance by comparing the pixel change-based and optical flow-based methods [65]. In addition, after manually selecting the shoulder boundary region as the RoI, Shao et al. [66] detected the movement of the region by calculating the differential in the longitudinal direction. A multi-task learning method for skin region segmentation was studied by Jorge et al., who refined various features extracted from a learning convolutional neural network (CNN) with BSS techniques, such as principal component analysis and ICA [67]. Brieva et al. also measured respiration through motion analysis, and other studies along these lines were conducted [68]. Although these motion analysis-based respiration measurement methods showed promise through experiments, they still involve two problems. The first is that no automated technology was applied to reliably detect RoIs to measure respiration. Failure to automatically detect a stable RoI not only implies a need to manually assign the RoI, but also requires extremely restricted movement so that the subject cannot deviate from the RoI. These methods are exceedingly difficult to apply to general-purpose measurement scenarios because the RoI must be periodically designated for continuous measurements when a subject moves. Although some studies estimated the area where movement due to respiration occurs as an RoI using the characteristics of respiration movement, this method is very vulnerable to motion noise, and its detection performance is thus unstable in environments with few motions, such as mobile and kiosk settings. In addition, Lin et al.'s approach was also used to estimate a chest RoI based on the face area, but the exact area may not always be detected depending on the subject's posture [69,70]. An inaccurate RoI causes noise in the background to be included in the analysis target, and thus degrades the respiration measurement performance, which makes the respiration measurement technology difficult to apply without accurate RoI detection technology, which tends to be available only in limited environments. Second, while most studies demonstrated promising performance, most of these experiments did not consider real-world situations with considerable noise elements [71]. The movement of the hand covering the RoI is larger and stronger than the fine movements of breathing, and distinguishing between these movements and breathing movements by means of pixel variation or optical flow is challenging. Therefore, such noise may cause large deformations in the estimation of the respiratory signal process and may significantly reduce the accuracy of the respiration measurement. To enable respiration measurement technology to be widely adopted, the respiration measurement process must be automated, and the respiration measurement performance must be guaranteed even in a general environment. Therefore, in this study, we overcome these two limitations by automating RoI detection through the application of deep learning-based techniques and analyzing respiratory movements using CNN models trained to be robust to noise.

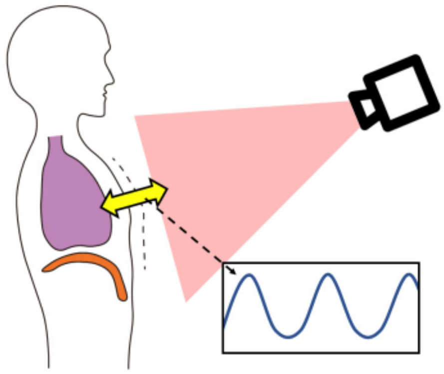

Figure 1. Principle of motion-based respiration measurement.

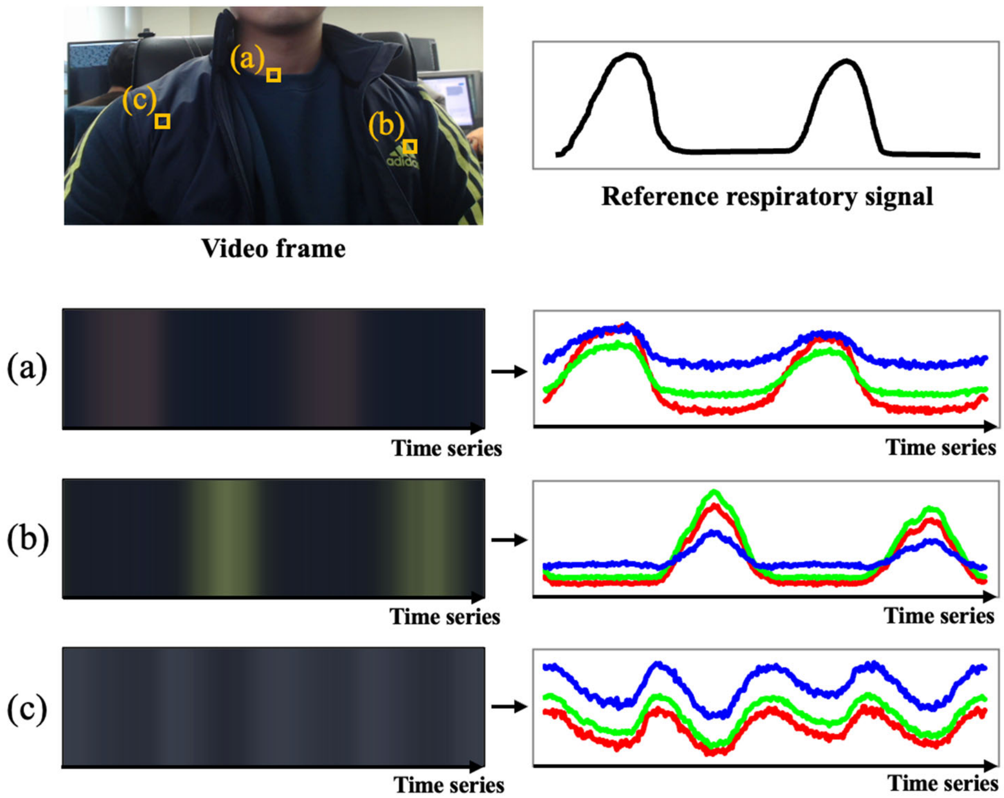

Figure 2. Examples of various pixel value changes that can occur as a result of the same breathing movement: (**a**) changes in pixel value that are similar to a reference breathing signal, (**b**) changes in pixel value that are opposite to the reference breathing signal, and (**c**) pixel value if the changes in the reference respiratory signal are not similar.

3. Materials and Methods

The entire process of the proposed method is illustrated in Figure 3. The RoI is detected in video frames using deep learning, and noise-robust breathing movements are detected from the detected RoI region. The detected motion information is accumulated for a predetermined time, and a respiratory signal is detected by purifying the accumulated signals to estimate a subject's rate of respiration. Figure 3a is described in detail in Section 3.1, Figure 3b in Section 3.2, Figure 3c in Section 3.4, and Figure 3d in Section 3.5.

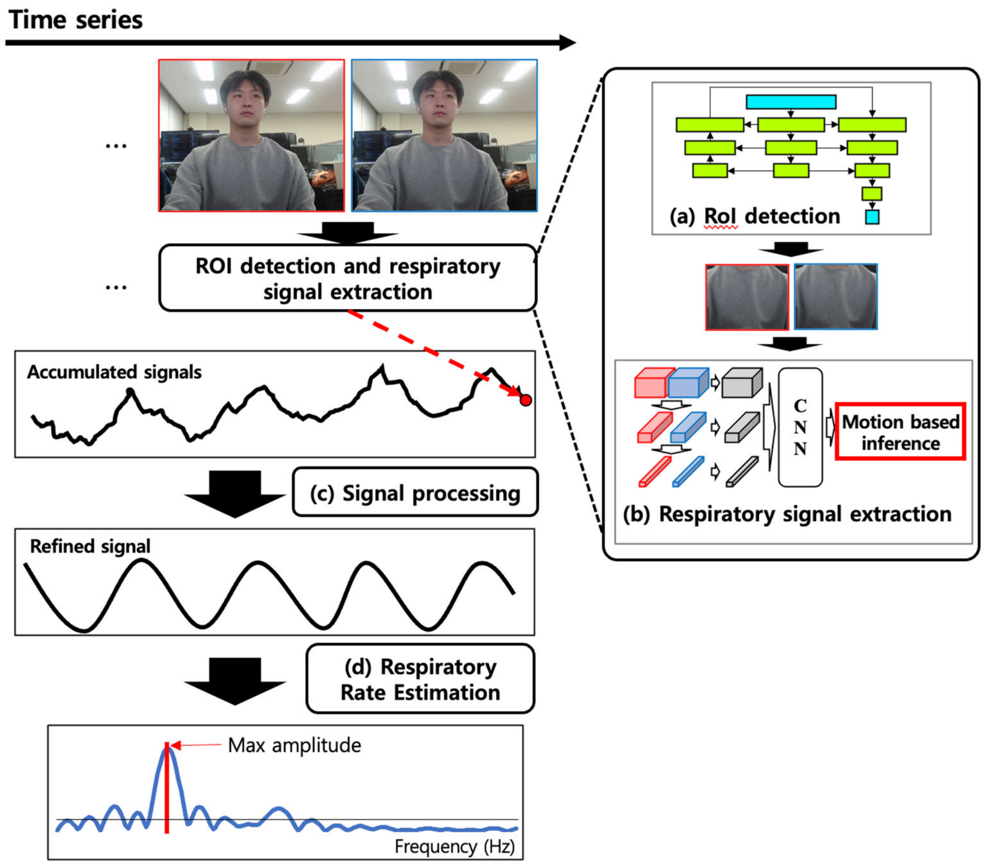

Figure 3. Overall process of the proposed method.

3.1. RoI Detection

In the proposed approach, we adopt a method to detect the location of specific body parts (face, torso, etc.) within an image where changes caused by respiration can be observed in order to automatically detect RoIs. In this section, we discuss the existing methods and their limitations in detail, along with some directions to improve their performance. In this section, we describe the proposed methods to improve existing RoI detection techniques by focusing on stably detecting the chest region as the RoI, because movement caused by respiration is most visible in the rising and falling of the chest [72]. To perform stable detection, a distinct feature is needed to identify the difference between the object to be detected and other objects. Face detection is relatively easy owing to the biologically distinct characteristics of human facial features [73], but few studies considered this approach because the characteristics of the chest area can vary depending on the pattern of clothing worn. Previous studies attempted to detect RoI by simply assuming the area below the face to be the chest area after detecting the face, but this is highly likely to select an incorrect area owing to factors such as the subject's posture. Hence, estimating respiratory motion information properly is difficult if only a portion of the chest region is included in the RoI, and stable measurement may not be possible due to background noise. To overcome these limitations to achieve continuous and stable chest region detection, we adopt a technique to estimate the main landmark location of the human body to detect RoIs.

Recently, technology to detect landmarks in the human body made significant progress [74]; in particular, Google announced BlazePose as a lightweight landmark detection technology that enables real-time operation even in mobile environments [75]. BlazePose detects the area of the face, which has the most distinctive characteristics in the human body, but little change in shape; then, it determines the area to detect landmark points based on the area of the face. Compared with existing body landmark detection technologies, BlazePose is designed to remove heatmap branches during inference and enables points to be directly predicted. Therefore, our proposed approach exhibits extremely fast performance in landmark detection on single objects and achieves a computational speed that enables real-time computation even in mobile environments with limited computing power. In addition, the proposed approach shows robust performance even with some points missing, such as when only the upper body appears in an image, by adding visibility to the inference results of the model. Considering that no significant difference in accuracy is observed compared to conventional methods, our proposed method uses landmark points detected on the human body through BlazePose to estimate RoIs. Because the chest area can be defined as the area between both shoulders and between the shoulder and pelvis, the chest RoI for respiration measurements can be defined in terms of a bounding box determined by the shoulder and pelvic points, as shown in Figure 4.

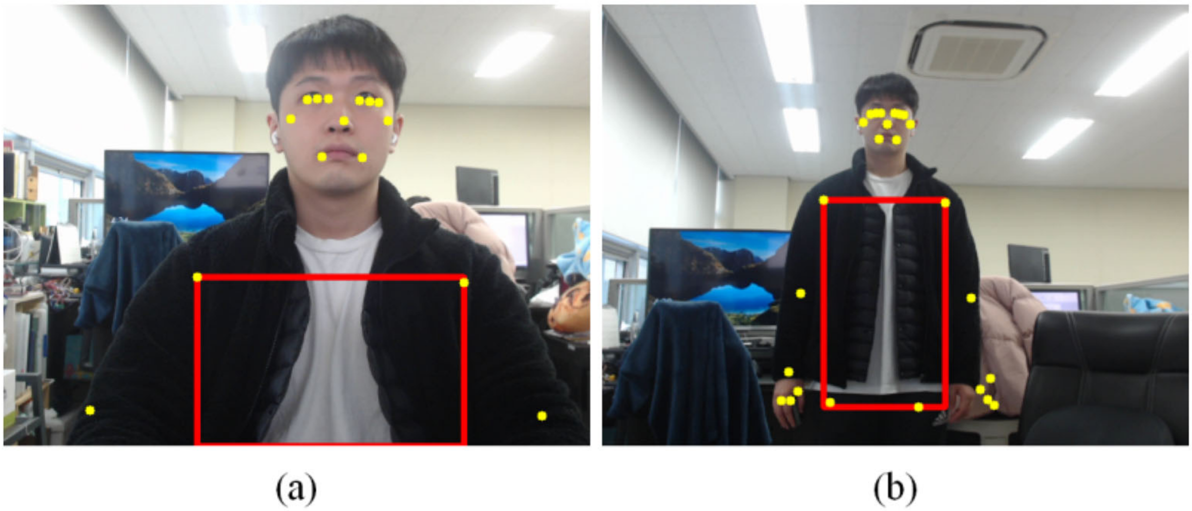

Figure 4. Example of chest RoI detection results based on body landmark point: (**a**) if only a portion of the chest area is in the image; (**b**) if all the chest area is in the image.

3.2. Motion-Based Respiration Measurement Method

We consider two methods to improve the stability of existing respiration measurement methods. The first is to suppress noise generated by movement not caused by breathing in the RoI. In an uncontrolled environment, the RoI of the measurement subject may include objects with other movements in addition to the chest and abdomen. Because breathing movements are relatively large, the movements of these objects must be suppressed so that the noise does not affect the motion signal. However, distinguishing breathing from noise based on the calculated amount of movement is difficult because the quantitatively measured amount of movement varies depending on the distance between the measured object and the camera, the amount of breathing, and the type of clothing worn. The second method is to discover additional characteristics that can be used for motion analysis. Most existing methods for estimating respiratory information by quantitatively measuring movements consider only vertical movement as indicating breathing. However, because respiration is a movement that causes expansion and contraction of the chest and abdomen, it is accompanied by movement in the left and right directions as well as in the vertical direction. In some cases, the left and right movements may not be observed due to breathing movements, or they appear larger, so a more stable measurement may be possible when left and right movement characteristics are considered. Therefore, to improve on the stability of existing methods, we consider distinguishing motion noise from fine motion caused by breathing and use additional features to measure motion.

In this study, we propose a learning-based CNN model that estimates motion through texture changes in the RoI. The structure of the proposed model is shown in Figure 5. The model extracts the features of an image through convolution (Conv1, Conv2, and Conv3 layers), which extracts the characteristics of the image by receiving RoI images of two adjacent video frames as input and calculates the difference in the features to obtain the feature maps (feature maps 1, 2, and 3). Further convolution then extracts expressive features in estimating motion information from these feature maps containing difference information between the two images and outputs a two-dimensional output that expresses the degree of exhalation and inhalation movement as values between −1 and 1. Because the output refers to the degree of motion of the two RoI image inputs, the average of the output is calculated to calculate the average degree of motion and is used as the average motion information. All convolutions, except convolution (Conv4-2), that output results include the ReLU activation function, and the Conv1, Conv2, and Conv3 layers include batch normalization. Although the model consists of convolution throughout the entire process, the size of the input image is fixed at 200 × 200 pixels for computational efficiency and characteristic normalization. Details of the parameters of the model are summarized in Table 2.

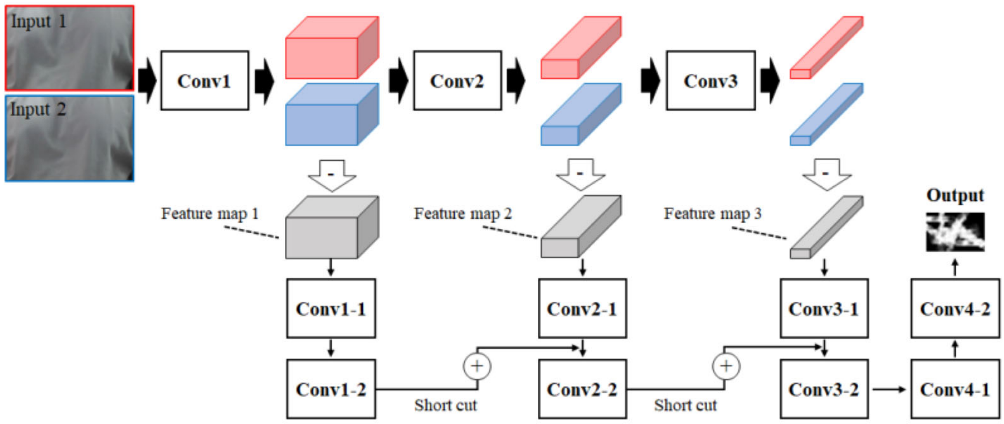

Figure 5. Proposed CNN architecture for respiration measurement.

Table 2. Model parameter of the proposed motion-based respiration measurement CNN.

Name	Kernel Size	Input Channels	Output Channels	Stride	Input Size	Output Size
Conv1	3 × 3	3	32	2	200 × 200	100 × 100
Conv1-1	3 × 3	32	32	1	100 × 100	100 × 100
Conv1-2	3 × 3	32	64	2	100 × 100	50 × 50
Conv2	3 × 3	32	64	2	100 × 100	50 × 50
Conv2-1	3 × 3	64	64	1	50 × 50	50 × 50
Conv2-2	3 × 3	64	128	2	50 × 50	50 × 50
Conv3	3 × 3	64	128	2	50 × 50	25 × 25
Conv3-1	3 × 3	128	128	1	25 × 25	25 × 25
Conv3-2	3 × 3	128	128	1	25 × 25	25 × 25
Conv4-1	3 × 3	128	128	1	25 × 25	25 × 25
Conv4-2	3 × 3	128	1	1	25 × 25	25 × 25

Because the initial layer of the CNN extracts low-level features, such as texture and edge, as the convolutions are repeated, the structural characteristics of the object are extracted. The model utilizes the complex structural characteristics of images [76]. However, the proposed method is not designed to use complex structural characteristics, but rather to estimate motion through changes in low-level features, such as texture. Therefore, relatively few layers and model parameters are needed. In addition, we adopt a shortcut structure so that the features extracted in the first layers can contribute to the last layer. This type of proposed model has several advantages.

First, the model can be designed as a lighter network to reduce the amount of computation required and enable faster and more efficient training. Deep learning-based respiration estimation shows excellent performance in terms of accuracy, but high-end devices are essential when real-time processing is required, owing to the large number of computational resources required. However, the proposed method can perform real-time calculations even in edge devices, such as mobile or embedded systems. In addition, the receptive field is also correspondingly small, which enables efficient calculation with less data. The receptive field can be defined as the size of the space of input neurons that can affect a single output neuron, which is defined as the size of the input image that can affect the output value [77]. The size of the receptive field is proportional to the extent to which context information can be utilized; a larger receptive field is useful for solving complex problems, but can easily lose detailed spatial characteristics of an image [78–80]. As the model has more contextual information available, more data with guaranteed diversity are required to prevent overfitting. The proposed model can be trained effectively owing to its lightweight design. In addition, the small receptive field of the model and the fact that it outputs motion information on the receptive field, which is a partial region of the image, in the form of a two-dimensional image, allows the information to be estimated independently. As the model has more contextual information, a sufficient dataset with guaranteed diversity is required to prevent overfitting. Because the proposed method has a small receptive field, it can learn intensively from the texture information of a local area of an image; thus, efficient augmentation can be achieved even with relatively little data.

Second, the existing methods use only movement in a predetermined specific direction as respiration information after calculating the motion information. However, because the proposed model calculates the amount of movement based on the change in the image without making a specific hypothesis on the breathing movement, the result can be derived by using the overall elements present in the image. In addition, this characteristic of the proposed method, which can handle more information, can also help to distinguish noise information included in the RoI.

3.3. Model Training

Data preprocessing: For training, time-sequential frames and data labeled with information on motion due to respiration are needed to estimate the degree of motion. In this study, we generated a dataset for respiration estimation, as shown in Figure 6, using video and respiration signals collected with a webcam and contact sensor. First, two time-adjacent frames of a captured video were used as input images, and the difference between the normalized respiration signals corresponding to the two images was obtained. The difference between the respiration signals was a value between -1 and 1, and the stronger the inhalation, the closer to 1, while the stronger the exhalation, the closer to -1. By multiplying the predefined body mask for the image with the movement value due to respiration, the torso area pixel values are assigned a movement value due to respiration, and the pixel values of areas other than the mask are assigned a value of 0 to create a respiration signal mask. The proposed model is trained to infer the torso mask for movement changes due to breathing by taking two images as the input.

Implementation details: The L_1 loss, as given in Equation (1), was used as the loss function, the Adam optimizer was used, and a total of 1,000,000 iterations with a batch size of 10. The learning rate started at 0.0001, and a step scheduler was applied to reduce the learning rate by half every 200,000 iterations.

$$\mathrm{L}_1 = \frac{1}{n}\sum_{i=1}^{n}|y_i - f(x_i)| \tag{1}$$

In the model inference, the size of the input image was fixed at 200×200 pixels, but we allowed the actual size and aspect ratio of the RoI to differ from the size of the input image. In this case, the characteristics of the image may be modulated by resizing to match the RoI image with a fixed image size. Therefore, we selected a bounding box of random location, size, and aspect ratio to consider these factors, and resized the bounding box to 200×200 pixels for use as input data. The size and aspect ratio of the bounding box were randomly selected from 0.5 to 1.5 times based on 200×200 at each learning time. In addition, color augmentation was performed by scaling and shifting randomly for each RGB channel of the image to include information on the texture changes of diverse colors. Furthermore, the proposed method needs 759,760,000(0.7 G) floating point operations (FLOPs) to respiration signal estimation.

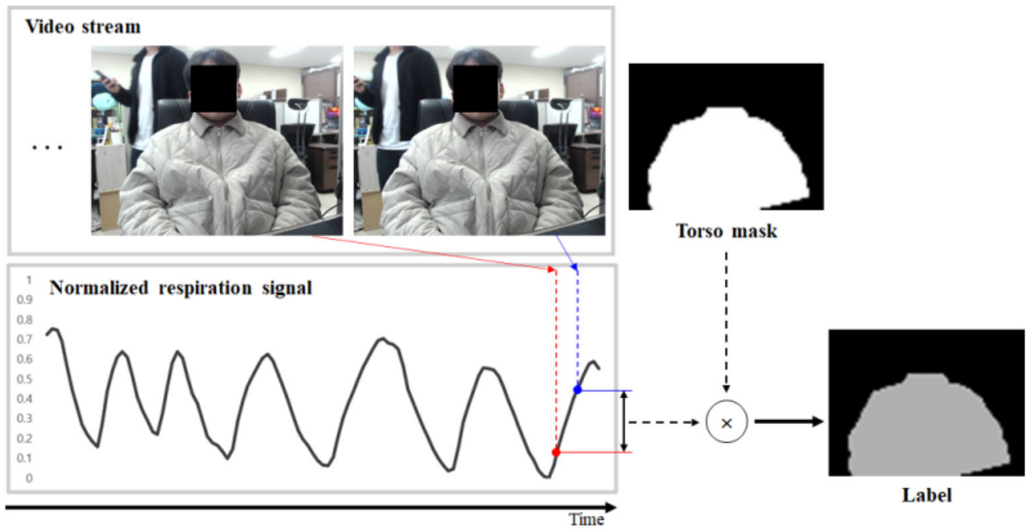

Figure 6. Constructing a training dataset construction from video and respiration signals with a contact sensor.

3.4. Refining Motion Estimation Model's Respiratory Signals

A time-series motion signal may be obtained by estimating the average motion information over time, as shown in Figure 7. Because the motion estimation model is trained to suppress noise components and detect only motion caused by respiration, a respiration signal that does not contain noise caused by movement other than breathing can be extracted. However, because the proposed motion estimation model compares the amount of motion between adjacent frames without reference, the accumulation of errors in each frame can cause bias in the detected motion signal, which can be considered as noise. Therefore, a post-processing operation must be performed to estimate respiration information by removing the corresponding noise component. In general, frequency-based post-processing methods are commonly considered for signal purification. These methods leave only specific components by using a band-pass filter to allow only the target frequency to pass through the acquired raw signal and to refine various signals, such as heart rate data. Because the deflection that occurs in the signal is a low-frequency component that exists in the entire signal, the trend of the respiratory signal can be removed by suppressing the low-frequency component. The low-frequency components of the signal can be obtained or removed through various methods, such as discrete Fourier transform (DFT) and moving average filtering-based methods. However, because the band-pass filter removes the defined specific frequency components, the pause between breaths is removed with low-frequency components, as shown in Figure 7a. The proposed method uses moving average filtering to obtain the signal. Considering that rates of 10–40 bpm are considered normal breathing, the window size of the filter used to obtain the low-frequency component was 6 s long, which may include the maximum period of breathing. Figure 7 shows an example of the raw signal where the trend occurred, and the breathing signal that removed the trend.

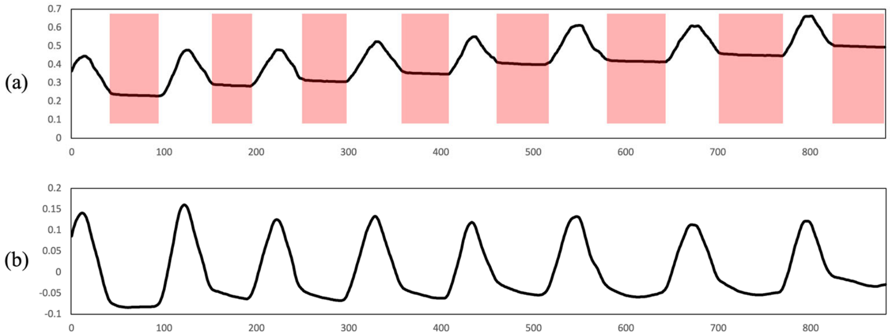

Figure 7. Examples of trended raw signals and detrending results: (**a**) raw signals, (**b**) detrending results.

3.5. Estimation of Respiratory Rate

Two methods are used to detect respiratory rate: a method based on peak-to-peak interval (PPI) and a method based on frequency analysis. PPI is the distance between adjacent peaks of the signal, which is the period of the signal. The respiratory rate (RR) may be calculated from PPI using Equation (2) in units of seconds.

$$RR = \frac{1}{PPI} \times 60 \tag{2}$$

The method based on frequency analysis calculates the frequency component of the signal, assumes that the frequency component with the largest amplitude is the respiratory frequency, and estimates the respiratory rate through the corresponding frequency. The respiratory rate can be calculated from the maximum frequency component using the following Equation (3):

$$RR = f_{max} \times 60 \tag{3}$$

To estimate the respiratory rate based on PPI, accurate peak detection must be performed. Because inaccurate peaks can significantly reduce the accuracy of respiration rate estimation, frequency analysis-based estimation methods may be suitable in real-world environments where stable peak detection may be difficult. Therefore, in this study, we conducted an experiment in which we estimated respiratory rates with a frequency analysis method using DFT.

3.6. Data Acquisition

All experiments were conducted using a notebook computer with an Intel Core i7-9750 processor (2.60 GHz), 16 GB RAM, and an NVIDIA GeForce RTX 2080 GPU, using the Windows 10 operating system. The image was acquired at 640 × 480 resolution using a Logitech C920 Pro webcam, as shown in Figure 8a. To prevent loss due to video compression, all captured images were stored as RGB raw data and photographed at a speed of 20 fps. The Vernier Go Direct Respiration Belt was used as reference equipment for breathing signals, as shown in Figure 8b, and data were acquired at 20 fps, according to the speed of the camera. All implementations required for the experiment were executed in the Python programming language (version 3.8), and we used libraries such as OpenCV, Pytorch, and ONNX Runtime to process images and to train and evaluate the models.

Figure 8. Data acquisition equipment for remote respiration measurement: (**a**) Logitech C920 pro, (**b**) Go Direct Respiration Belt.

The respiration dataset was obtained from 14 participants. The subjects breathed comfortably approximately 1 m from the camera, and the front of their upper body was photographed by the camera while they were guided to breathe according to a provided breathing guideline signal. The guidelines were used to induce three types of respiratory patterns to include various respiratory patterns in the data, and images and respiratory signals were photographed for 260 s per the guidelines. Guideline #1 induced breathing at 40, 35, 30, 25, 20, 15, 10, 15, 20, 25, 30, 35, and 40 bpm for 20 s, and Guideline #2 induced breathing in the form of rapid change in bpm speed to reflect rapid changes in the data. Guideline #3 induced breathing that included pauses during breathing to reflect a stationary state in the learning data. To include movement information other than breathing in each image, moving objects were included in the background, in addition to the subjects, as shown in Figure 9. For each subust, 15,700 image frames were photographed, 15,600, 5200 for each guideline, and a total of 218,400 image frames and respiratory signal values corresponding to those frames were obtained for all subjects.

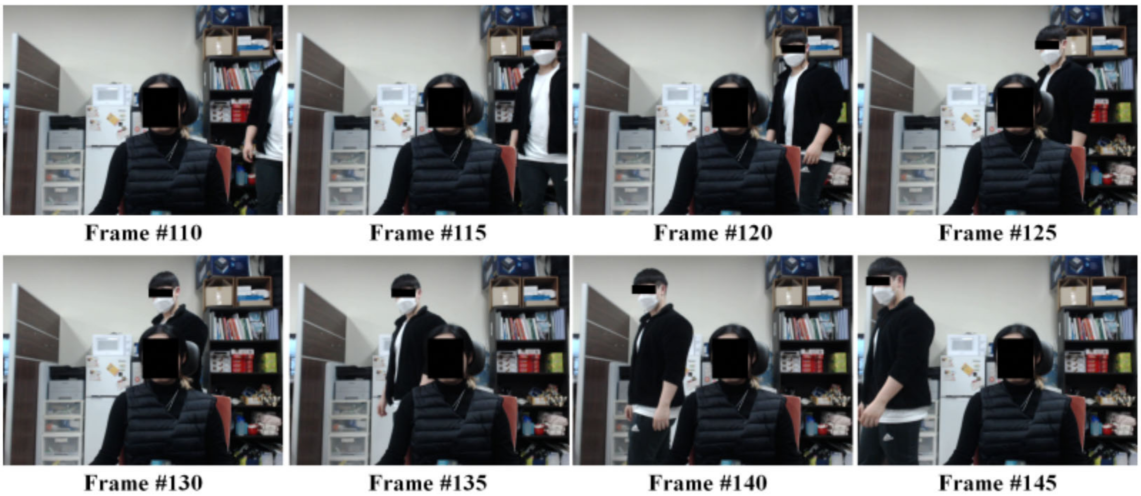

Figure 9. Example of respiration dataset with motion and background noise.

We obtained testing data to evaluate the performance of the model for 13 people in a spontaneous breathing situation without any separate control related to the subjects' breathing patterns. The subjects were photographed at a close distance of about 0.5 m and a long distance of about 1.5 m and were also photographed during thoroughly controlled movement and a situation in which movement noise was generated in RoI. For the situation with noise, the subjects' RoI was photographed for an environment in which various objects, such their hands and clothes, were also intermittently moving.

3.7. Performance Validation and Metrics

We evaluated the accuracy of the proposed method in comparison with a reference respiratory rate. The respiratory rate must accumulate data for a specific time, after which a new bpm can be calculated for each new input frame. Therefore, in this study, based on the experimental results of Section 4.1, after accumulating data for 17 s, the bpm for all subsequent frames was calculated using the sliding window method. The following analysis methods were used to verify the respiration measurement performance. To verify the accuracy of the estimated respiratory rate, the mean absolute error (MAE) was used to determine the absolute difference between the two measurements (reference and measured bpm), as shown in Equation (4).

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\mathrm{RR}^{i}_{pred} - \mathrm{RR}^{i}_{gt}\right| \tag{4}$$

The Pearson correlation coefficient (r) between the two signals was used to determine the similarity between the estimated signal and the estimated respiratory rate to the original signal. This metric evaluates the correlation between two data points and can be used as an indicator to evaluate the similarity of the trends of the two signals. The following is implied by r: the closer r is to 1, the more positively the two data samples are correlated; the closer to −1, the more negatively the two samples are correlated; and the closer to 0, the less correlated the samples are. This value can be calculated from Equation (5) for both respiratory signals, and we used the average as an indicator to determine the quality of the estimated respiratory signals.

$$\mathrm{r} = \frac{\sum_{i=1}^{n}\left(s_{gt}^{i} - \overline{s_{gt}}\right)\left(s_{pred}^{i} - \overline{s_{pred}}\right)}{\sqrt{\sum_{i=1}^{n}\left(s_{gt}^{i} - \overline{s_{gt}}\right)^{2}}\sqrt{\sum_{i=1}^{n}\left(s_{pred}^{i} - \overline{s_{pred}}\right)^{2}}} \tag{5}$$

The intraclass correlation coefficient (ICC) was also calculated to analyze the degree of agreement between the reference bpm and bpm measured using the contactless respiration measurement method. ICC was calculated using Equations (6)–(8); the closer to 0, the lower the consistency, and vice versa.

$$\overline{\mathrm{RR}} = \frac{1}{2n}\sum_{i=1}^{n}\left(\mathrm{RR}_{gt}^{i} + \mathrm{RR}_{pred}^{i}\right) \tag{6}$$

$$s^{2} = \frac{1}{2n}\left\{\sum_{i=1}^{n}\left(\mathrm{RR}_{gt}^{i} - \overline{\mathrm{RR}}\right)^{2} + \sum_{i=1}^{n}\left(\mathrm{RR}_{pred}^{i} - \overline{\mathrm{RR}}\right)^{2}\right\} \tag{7}$$

$$\mathrm{ICC} = \frac{1}{ns^{2}}\sum_{i=1}^{n}\left(\mathrm{RR}_{gt}^{i} - \overline{\mathrm{RR}}\right)\left(\mathrm{RR}_{pred}^{i} - \overline{\mathrm{RR}}\right) \tag{8}$$

In addition, we used a Bland–Altman plot to analyze whether error bias was present along with the mean of different (MOD) and limit of agreement (LOA) indicators for further performance analysis. We performed comparisons with existing methods to evaluate the relative performance of the proposed method. Two methods were used for the comparison, including an intensity variation-based method for RGB pixels [57] and an optical flow-based method [62]. The method of detecting respiratory volume from images followed the existing method, but the RoI detection method followed the method proposed in this study to prevent performance differences due to differences in RoI detection accuracy.

4. Experimental Results

4.1. Analyze the Impact of Changing the Time Window Size

A respiratory signal of a predetermined length or greater is required to refine the respiratory signal and detect the respiratory rate. Short respiratory signals may not include many respiratory cycles, resulting in poor estimation accuracy, and excessively long respiratory signals may increase the time required for DFT computation to purify the signal and calculate the bpm. Therefore, we conducted an experiment to select the optimal window size considering accuracy and inference speed. The experiment derived the results through 5-fold cross-validation for the training data. Table 3 and Figure 10 show the changes in the accuracy of bpm estimation and processing time according to the changes in the size of the window. As a result of the experiment, the bpm estimation accuracy exhibited the highest performance when the window sizes were 16 and 17 s. Notably, when the window size increased to more than 18 s, the accuracy decreased. This can be advantageous for periodic analysis, including many breathing cycles; however, using a window size larger than a certain level increases the probability of bpm changing or including noise elements, which makes accurately estimating bpm difficult. Therefore, choosing an optimal window size that is neither excessively small nor large is important for ideal respiration estimation. In terms of computation time, when the window size was less than 5 s, the computation time was noticeably short, but the difference in processing speed between time windows of more than 10 s was approximately 0.1 ms, so the increase in window size was not disadvantageous in terms of processing speed. Therefore, in this study, 17 s was used as the optimal window size considering both bpm estimation accuracy and computation time.

Table 3. The changes in the accuracy and inference time according to window size.

Window Size (s)	MAE (bpm)	Inference Time (ms)	Window Size (s)	MAE (bpm)	Inference Time (ms)
1	5.5430	0.3314	16	0.1859	0.8533
2	3.2044	0.4635	17	0.1848	0.8882
3	1.8921	0.6130	18	0.2144	0.8594
4	1.2553	0.7679	19	0.2383	0.8625
5	0.9901	0.7876	20	0.2193	0.8385
6	0.8733	0.7889	21	0.2068	0.8442
7	0.6581	0.7971	22	0.2345	0.8951
8	0.5102	0.7986	23	0.2534	0.8564
9	0.4739	0.7920	24	0.2382	0.9244
10	0.4603	0.8154	25	0.2129	0.9412
11	0.3785	0.8468	26	0.2277	0.8854
12	0.3145	0.8501	27	0.2148	0.8739
13	0.2869	0.8260	28	0.2263	0.9089
14	0.2441	0.8311	29	0.2579	0.8530
15	0.2350	0.8423	30	0.2524	0.9211

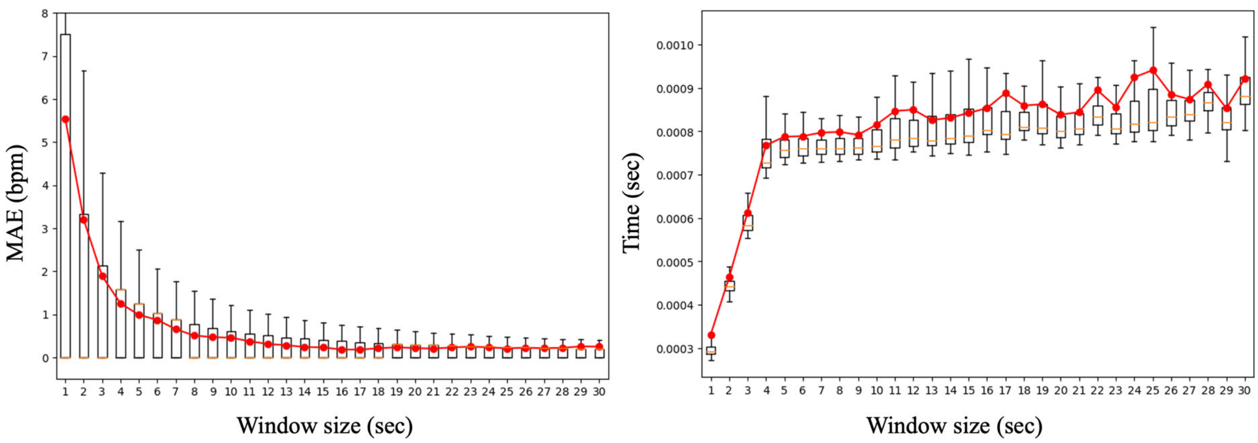

Figure 10. Comparison of respiration estimation accuracy and computational time according to window size change.

4.2. Results of Respiration Estimation in a Restricted Movement Environment

To verify the respiration measurement performance of our proposed method, we conducted an experiment to compare the results of estimating the respiratory signal and rate with the reference using test data. To confirm the respiration measurement performance in ideal situations where movement is strictly controlled, we conducted experiments only on scenes #1 and #2 of the test data. Details of the experimental results can be found in Tables 4 and 5, and the Bland–Altman plot results are shown in Figure 11. According to the experimental results, the MAE of the method proposed in scene #1 was the most accurate at 0.084, and that of the optical flow-based method was the most accurate at 0.123 in scene #2. When the data from scenes #1 and #2 were used, the optical flow-based method showed the highest performance at 0.107. We confirmed through ICC that the proposed method and the optical flow-based method exhibited excellent bpm estimation performance. Although the optical flow-based method showed the best performance in the experiment when movement was controlled, the difference was potentially not significant because bpm was calculated with a precision of a 0.36 level due to the nature of the frequency analysis method. Both methods showed the small error of a 0.1 bpm level and an exactly accurate respiration measurement was possible. Furthermore, a slight degradation in performance was observed at a greater distance than at a closer distance, which can be attributed to the greater influence of optical noise with increasing distance. However, even at a distance of 1.5 m, the respiration measurement methods exhibited good accuracy. The proposed method and optical flow-based method recorded small MODs and LOAs in all situations, which means that the error bias was small and the agreement between the reference and the data estimated through the two methods was remarkably high. The Bland–Altman plot in Figure 11 shows that the two methods had little error bias, and the agreement between the data was also high. Each point in the plot is distributed discretely by bpm calculation precision, and the density of the overlapped points is shown in color (red indicates higher density and blue lower density).

Table 4. Respiration measurements in motion-controlled environments.

Scene #	Method	MAE (bpm)	MOD (bpm)	LOA (bpm)	ICC	Mean of
#1	Ours	0.08 ± 0.14	-0.00	± 0.38	0.999	0.93 ± 0.04
#1	Optical flow [62]	0.09 ± 0.16	-0.01	± 0.36	0.999	0.95 ± 0.03
#1	Intensity [57]	3.33 ± 3.30	0.95	± 11.06	0.326	0.49 ± 0.26
#2	Ours	0.16 ± 0.16	0.04	± 0.56	0.999	0.93 ± 0.03
#2	Optical flow [62]	0.12 ± 0.15	0.05	± 0.71	0.999	0.94 ± 0.04
#2	Intensity [57]	2.16 ± 2.69	1.00	± 6.40	0.518	0.55 ± 0.27
Total	Ours	0.12 ± 0.15	0.02	± 0.48	0.999	0.93 ± 0.04
Total	Optical flow [62]	0.11 ± 0.15	0.02	± 0.57	0.999	0.95 ± 0.04
Total	Intensity [57]	2.75 ± 3.03	0.97	± 9.03	0.429	0.52 ± 0.27

Table 5. Analysis of error statistics in a motion-controlled environment.

Scene #	Method	*p*-Value
#1	Ours–optical flow [62]	0.599
#1	Ours–intensity [57]	0.000
#2	Ours–optical flow [62]	0.531
#2	Ours–intensity [57]	0.000
Total	Ours–optical flow [62]	0.928
Total	Ours–intensity [57]	0.000

Table 5 shows the results of the statistical verification of the difference between the bpm error group of the proposed method and the bpm error group of the existing methods through a *t*-test, and there was no statistically significant difference between the optical flow-based method and the proposed method for all motion-controlled situations. Intensity-based methods recorded low performance values in all the experimental results. This can be attributed to the limitations of the pixel variation-based method mentioned in Section 2.2. Figure 12 illustrates this problem by comparing the estimated signals. Because the reference signal used in the experiment was acquired through a belt-type respiratory measurement sensor, the detailed shape of the waveform may vary depending on the degree of looseness of the belt and the position in which it was worn. Therefore, in this study, the figure of correlation, which can confirm the similarity between the estimated signal and the reference, is not used as an absolute performance indicator, but rather as data to check whether the estimated signal represents a breathing pattern well. From a similarity perspective, we can see that intensity-based methods do not express respiratory signals well with extremely low values, and that the proposed method and optical flow-based methods showed a high degree of agreement of 0.9 or more in all situations. In fact, it may be confirmed from Figure 12 that the signal estimated through the corresponding methods was almost similar to the reference.

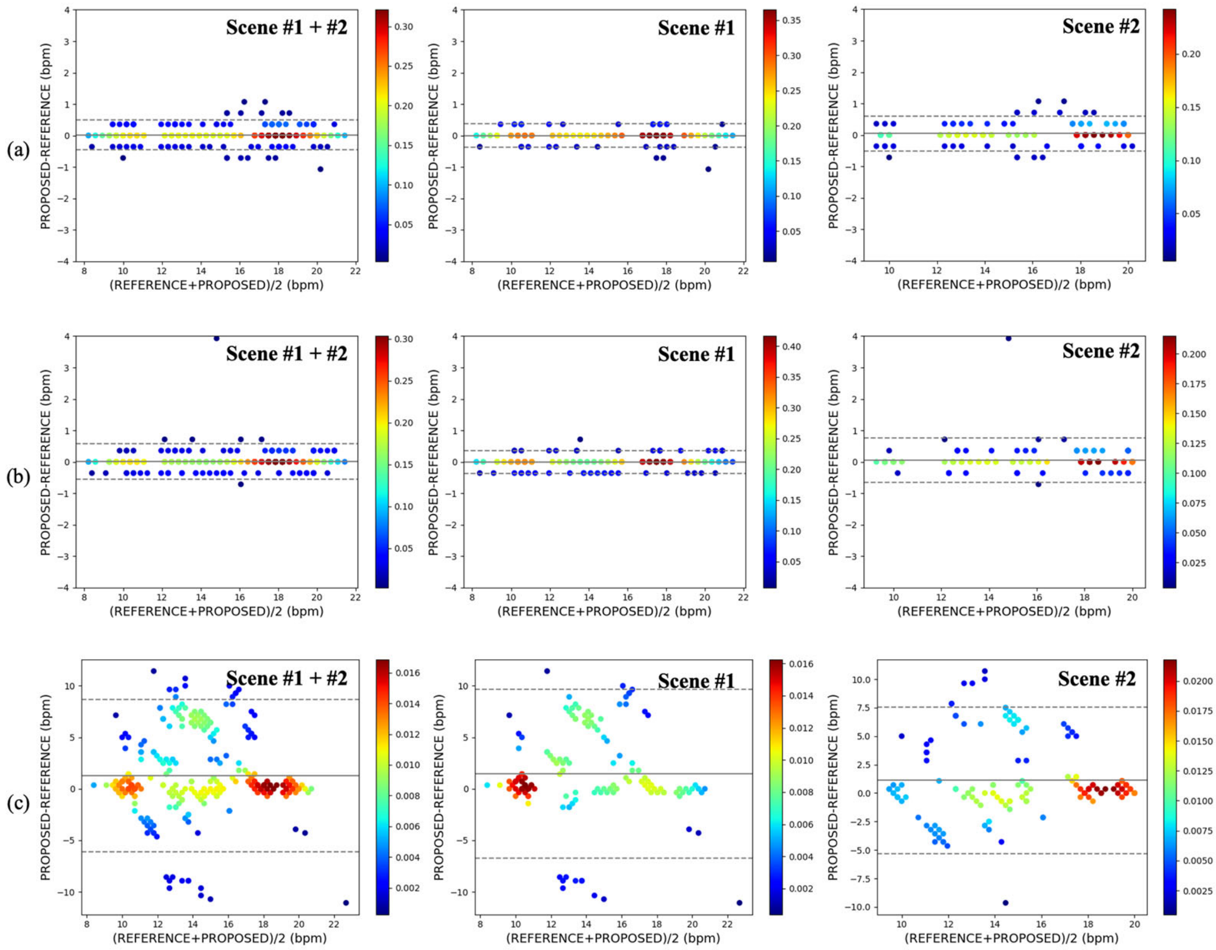

Figure 11. Bland–Altman plot of three breathing estimation methods for scene #1 and #2: (**a**) ours, (**b**) optical flow-based, and (**c**) intensity-based.

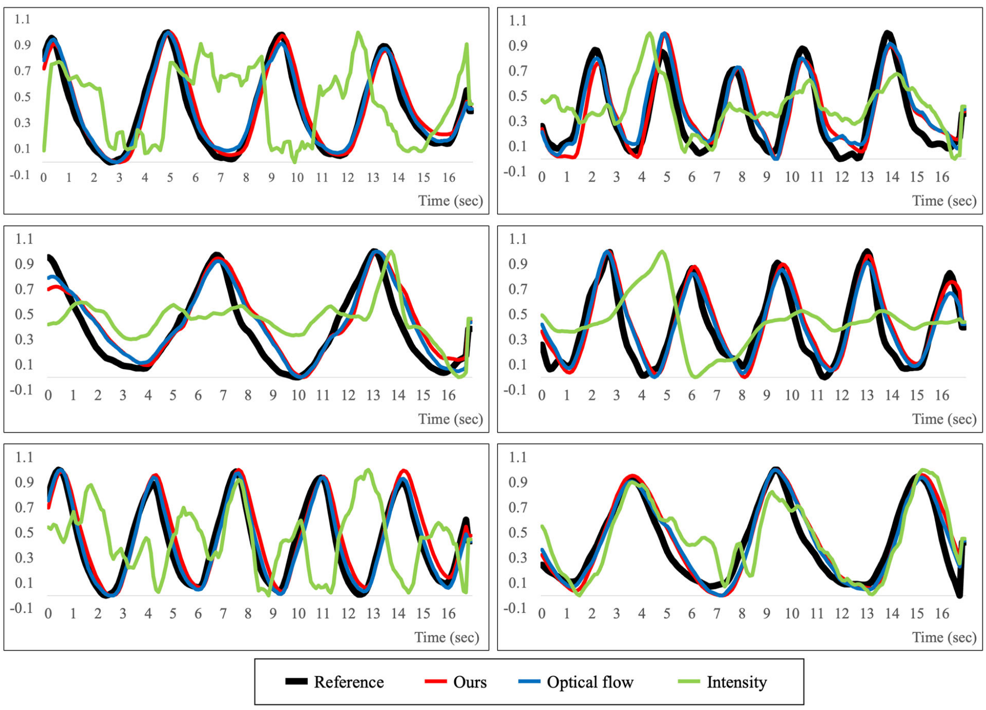

Figure 12. Comparison analysis of reference and inference respiration signal.

4.3. Estimation of Respiratory Rate in a Free-Motion Environment

To apply the proposed method in practice, stable performance in the event of motion noise is necessary, along with a thoroughly controlled environment. Therefore, we conducted an experiment on scenes #3 and #4, which were situations in which movements other than breathing occurred in the RoI. The experiment was conducted in the same manner as before, and the actual experimental results are shown in Tables 6 and 7 and Figure 13. The results show that the proposed method measured respiration accurately, and the values did not differ significantly from those obtained in the controlled environment. The optical flow-based method, which showed excellent respiration measurement performance in previous experiments, increased the overall MAE of the estimated bpm by a factor of more than 20, and the intensity-based method also decreased significantly. The correlation between the reference signal and the optical flow-based method also showed a low agreement of up to 0.4 in this experiment, in contrast to the high value of 0.9 or more obtained in the previous experiment. The ICC of the optical flow-based method was also quite low at 0.2 levels, unlike in previous experiments. The Bland–Altman plot in Figure 13 and the MOD and LOA results in Table 6 show that the error bias was large and the agreement between the estimated data was remarkably low. From the *t*-test results in Table 7, it may be confirmed that the proposed method shows a statistically significant error compared to all existing methods in a noisy environment.

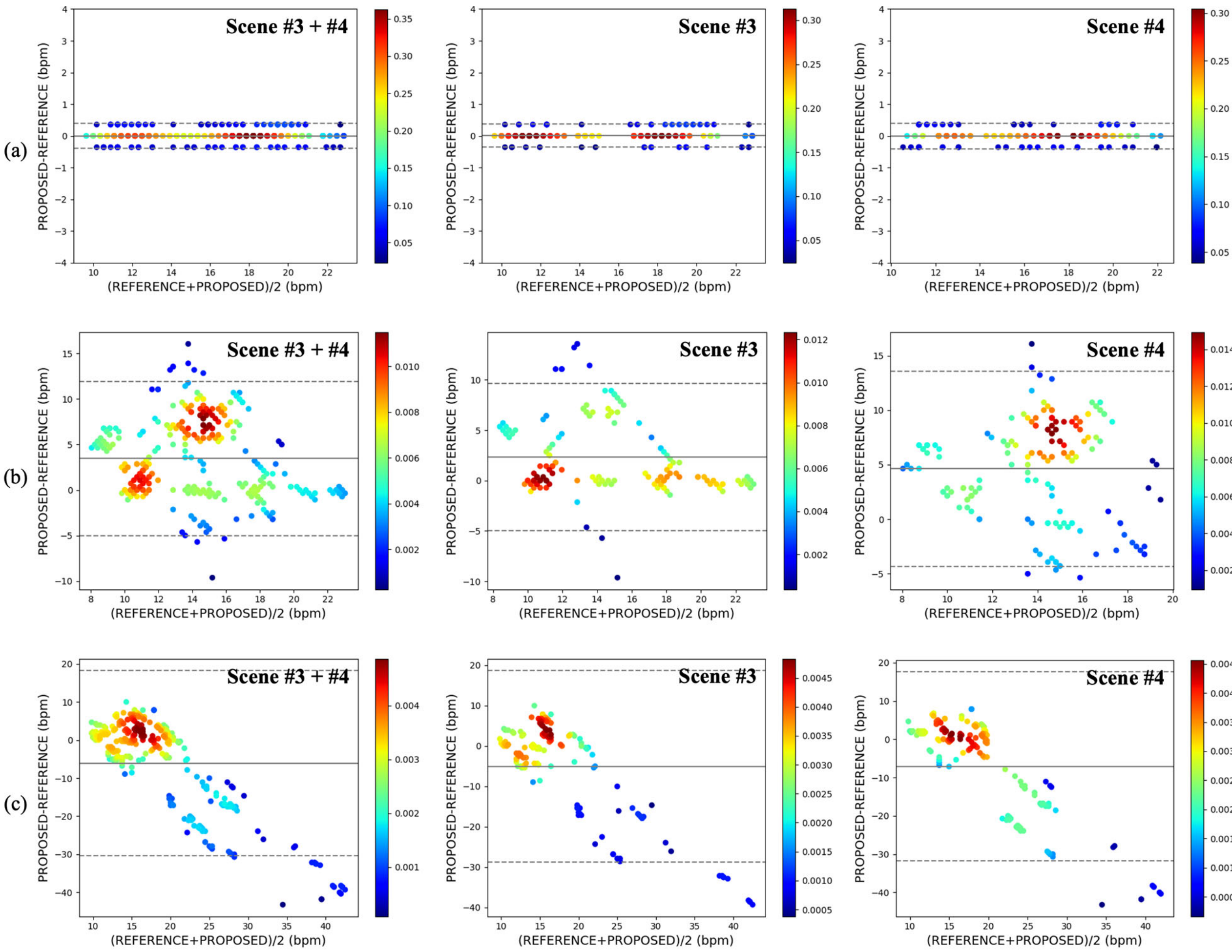

Figure 13. Bland-Altman plot of three respiration estimation methods for scene #3, #4: (**a**) ours, (**b**) optical flow-based, and (**c**) intensity-based.

Table 6. Respiration measurements in environments with motion noise.

Scene #	Method	MAE (bpm)	MOD (bpm)	LOA (bpm)	ICC	Mean of
#3	Ours	0.10 ± 0.15	0.01	±0.37	0.999	0.91 ± 0.03
#3	Optical flow [62]	2.84 ± 3.45	2.34	±7.32	0.521	0.44 ± 0.28
#3	Intensity [57]	8.33 ± 9.22	−5.01	±23.70	0.209	0.21 ± 0.15
#4	Ours	0.12 ± 0.15	−0.00	±0.41	0.999	0.92 ± 0.04
#4	Optical flow [62]	5.57 ± 3.33	4.63	±8.96	0.192	0.26 ± 0.20
#4	Intensity [57]	9.45 ± 8.37	−7.13	±24.69	0.237	0.22 ± 0.15
Total	Ours	0.11 ± 0.15	0.00	±0.39	0.999	0.92 ± 0.04
Total	Optical flow [62]	4.20 ± 3.64	3.48	±8.48	0.202	0.35 ± 0.26
Total	Intensity [57]	8.89 ± 8.81	−6.07	±24.29	0.134	0.22 ± 0.15

Table 7. Analysis of error statistics in environment with motion noise.

Scene #	Method	*p*-Value
#1	Ours–optical flow [62]	0.000
#1	Ours–intensity [57]	0.000
#2	Ours–optical flow [62]	0.000
#2	Ours–intensity [57]	0.000
Total	Ours–optical flow [62]	0.000
Total	Ours–intensity [57]	0.000

Table 8 shows the results of analyzing the statistical difference between the error group of bpm measured in the motion noise environment and the error group of bpm measured in the restricted motion environment for each noncontact respiration measurement method. Although the proposed method did not show statistically significant errors in the two situations, the results confirm that the other two methods showed statistically significant errors. Table 8 shows the results of analyzing the statistical difference between the error group of bpm measured in the motion noise environment and the error group of bpm measured in the restricted motion environment for each noncontact respiration measurement method. Notably, the proposed method did not exhibit statistically significant errors for the two situations, and the other two methods exhibited statistically significant errors.

Table 8. Analysis of error statistics between estimation with motion noise and estimation without motion noise.

Distance (m)	Method	*p*-Value
0.5 m	Ours	0.593
0.5 m	Optical flow [62]	0.000
0.5 m	Intensity [57]	0.000
1.5 m	Ours	0.999
1.5 m	Optical flow [62]	0.000
1.5 m	Intensity [57]	0.000
Total	Ours	0.715
Total	Optical flow [62]	0.000
Total	Intensity [57]	0.000

Figure 14 shows that reference and model-based breathing signals maintained stable waveforms at the time of motion noise, whereas optical flow and intensity-based methods showed significant damage to signal waveforms. Although the proposed method uses an image-based inference model that can respond to motion noise when it is included and learns from data containing information about motion noise, existing methods are vulnerable to motion noise. This experiment shows that the proposed method is much more robust to motion noise than conventional techniques. Hence, the results show that the proposed method can be applied to real-world environments where noise can occur with much higher reliability than conventional methods.

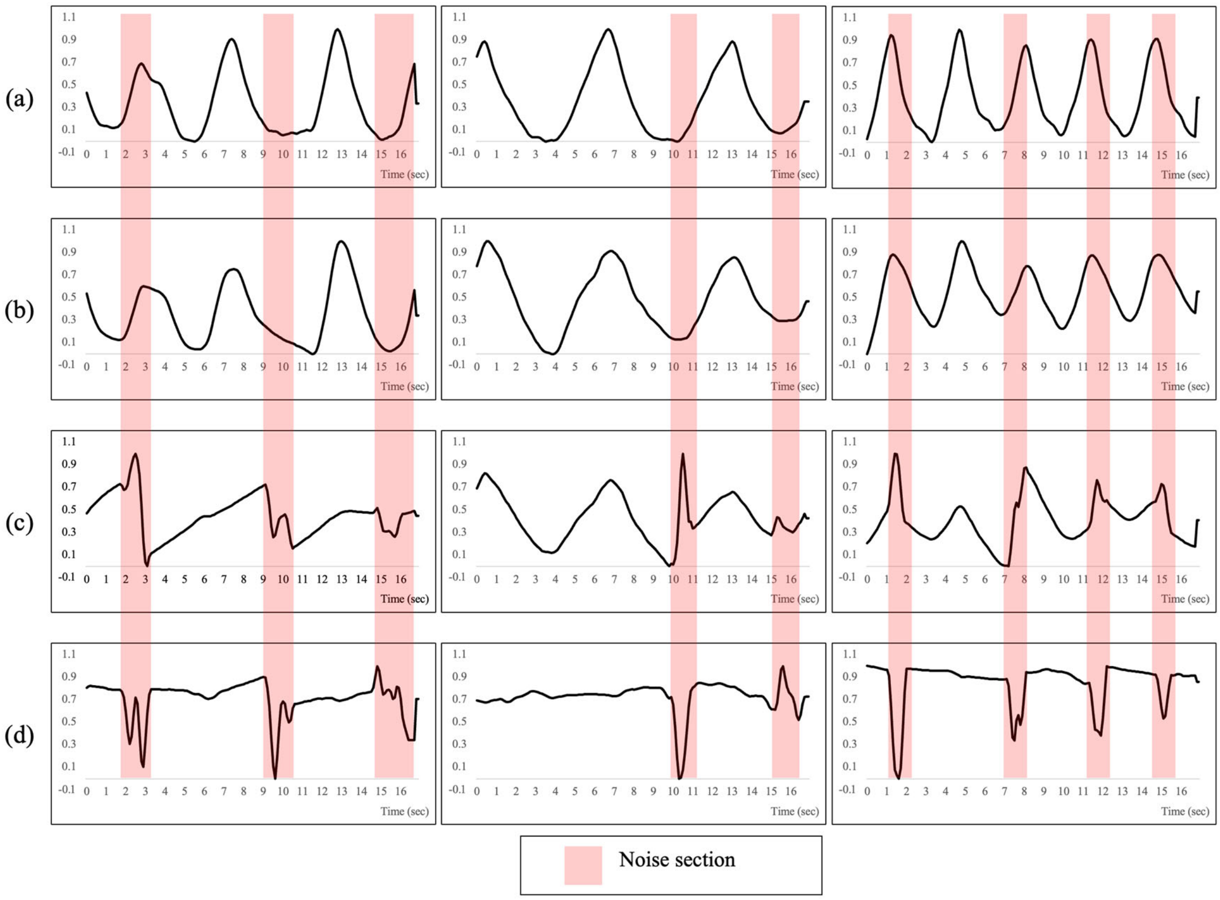

Figure 14. Examples of references and estimated signals on motion noise environment: (**a**) reference, (**b**) ours, (**c**) optical flow-based, and (**d**) intensity-based.

4.4. Inference Speed

The goal of the proposed method is to measure respiration in real time using standard webcams and computing devices without using contact devices. Therefore, an experiment was conducted to measure the computational time required to perform the entire process using the proposed method. To measure the calculation time, the execution time of each step was measured for real-time operation with an actual webcam. The measurement process was repeated a total of 10,000 times. In Table 9, among the steps for respiratory detection, RoI detection, signal purification, and bpm calculation were performed using only the CPU, and respiratory movement detection through the model was the result of using a GPU. Among the entire process for estimating respiration, the process of detecting the torso area with RoI using BlazePose was the longest at 19.57 ms. Other computational processes showed a short time of less than 3 ms. The total processing time, including the RoI detection process, is 21.87 ms, achieving real-time at an arithmetic speed of approximately 45 fps. However, GPUs are difficult to consider as typical computing devices. To discuss more general-purpose utilization, in this study, the speed at which GPUs were not used and only CPU resources were used was also verified through experiments. Table 10 shows the results of performing the operation of the entire process using only the CPU without using the GPU. The average performance speed of the motion measurement step, which performed the operation using the GPU, increased by about 3 ms from 1.96 ms to 4.58 ms. In addition, as the computational load of the CPU increased, the execution time of RoI detection increased by about 5 ms, signal refinement by about 0.04 ms, and RR estimation by about 0.09 ms. The total processing time in this case is 29.5 ms, which is a processing speed of about 34 fps. Given that the standard for commonly used real-time operations is 30 fps, these results show that the proposed method can perform real-time processing even with general computing and webcam hardware.

Table 9. Inference time of each respiration estimation on GPU.

Task	RoI Detection	Motion Measurement	Signal Refinement	RR Estimation	Total
Time(ms)	19.57	1.96	0.13	0.20	21.87

Table 10. Inference time of each respiration estimation on CPU.

Task	RoI Detection	Motion Measurement	Signal Refinement	RR Estimation	Total
Time(ms)	24.45	4.58	0.17	0.29	29.50

5. Discussion and Limitations

Recent studies on noncontact respiration measurement were widely adopted following the interest of noncontact heart rate measurement. Nevertheless, to the best of our knowledge, the proposed approach is the first to measure respiration in a fully automated manner, and previous studies needed a pre-defined torso RoI for respiration measurements and only developed an advanced respiratory signal extraction algorithm. The proposed method detects up to the torso area and extracts respiratory signals from the detected RoI. In addition, our proposed respiration measurements have a processing speed of less than 30 ms, even in a CPU environment, which generalizes noncontact respiration measurement and presents new directions for applications in various fields. In addition to complete automation of remote respiratory rate measurements, the results of several experiments verified that noise removal, which was is not solved by existing methods, can be successfully performed through deep learning methods that infer the rate of respiration rate in motion noise environments. Therefore, the proposed method is designed to be reliable and robust to motion noise.

The proposed method can be widely adopted and can also be executed on inexpensive and commonly used webcams and laptops, in contrast to existing methods. However, our proposed method is an early version of the real-time automatic respiration measurement technology with model-based motion analysis technology, and continuous improvement remains needed to develop the techniques that make up the proposed method further. First limitation: Our approach can be expanded to simultaneously measure multiple breaths in a single image. Moreover, there is a need for the optimization and verification of the various structures and scales that the proposed CNN model can process. Second limitation: Further experiments can be conducted by acquiring more data to improve the performance of the model and additional augmentation can be applied. In addition, it is necessary to verify the performance of various factors, such as measurement distance, measurement posture, color, and type of clothing worn by the subject, and additional experiments can be conducted on changes in the movement of objects covering the RoI. In addition, because the frequency analysis-based method used to estimate the respiratory rate uses the strongest frequency component as the respiratory rate at a window of 17 s, reflecting changes in respiration may be difficult if a subject's respiratory pattern changes rapidly. Because the proposed method can estimate stable respiratory signals as well as measure the respiratory rate, it can be extended to detect additional information (depth of respiration, apnea interval, etc.) that can be estimated from respiratory signals. Third limitation: The respiration signal extraction is affected according to the ROI selection method. The proposed method uses the body detector to measure respiration in torso ROI, so the measurement results may vary depending on the performance of the body detector. For this, a one-stage respiration measurement method combining ROI selection and respiratory rate measurement can be proposed, and the one-stage respiration measurement model structure for this can be referred to the following literature [81,82]. If these limitations are addressed, the proposed method can be applied to a variety of applications, including medical, forensic investigation, fitness, and emotional estimation, and enhancing its performance for further applications can be considered by combining it with technology for estimating the heart rate from skin color.

6. Conclusions

In this study, we proposed a contactless breathing measurement method that can universally utilize standard RGB cameras. Existing contactless breathing measurement methods are not widely adopted because of the need for a separate device, difficulty in automatically detecting RoI, and limitations in that measurement is impossible when unintended movement occurs. In this study, it is shown that the entire process can be automated using only RGB cameras for respiration measurements and the technology to automatically detect RoI, and the existing limitations are improved by using a CNN-based method designed to detect respiratory movements with high resistance to noise. Segmentation technology and lightweight body landmark-based technology were presented as alternatives to the existing RoI detection technology, and an environment in which the two methods could be applied was considered.

For respiratory motion analysis, we proposed a model that learns the movement occurring in two RoIs by applying various augmentation techniques to improve its performance. We also presented a method for purifying noise components that may be included in a motion signal, and we applied a method to detect respiratory rates from the purified signal. The performance of the proposed method was compared with representative existing methods based on optical flow and pixel intensity. The proposed method enables accurate respiration measurement with an error of approximately 0.1 bpm and did not show performance degradation even in a noisy environment that notably degraded the performance up to 8 bpm errors in existing methods. It can also operate in real time about 30 FPS and need 0.7 GFLOPs for inference. This shows that it can be applied not only to a CPU environment but also to a mobile device. Our experimental results show that the proposed method can be implemented in various applications to overcome existing limitations. However, the problem of performance degradation due to distance still needs further investigation, and further analysis remains necessary because we analyzed the effect of the movement of the measurement target on performance, rather than that of movements occluding the RoI.

Author Contributions: Conceptualization, E.C.L.; methodology, E.C.L. and H.H.; validation, K.L.; investigation, H.H.; data curation, H.H.; writing—original draft preparation, K.L.; writing—review and editing, E.C.L. and K.L.; visualization, H.H.; supervision, E.C.L.; project administration, E.C.L.; funding acquisition, E.C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a 2020 Research Grant from Sangmyung University.

Institutional Review Board Statement: Based on the 13-1-3 of the Enforcement Regulations of the Act on Bioethics and Safety of the Republic of Korea, ethical review and approval were waived (IRB-SMU-S-2021-1-005) for this study by Sangmyung University Institutional Review Board, because this study uses only simple contact measuring equipment or observation equipment that does not follow physical changes.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The obtained data cannot be shared because it was agreed that it could be used only for this study.

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

References

1. Kevat, A.C.; Bullen, D.V.R.; Davis, P.G.; Kamlin, C.O.F. A Systematic Review of Novel Technology for Monitoring Infant and Newborn Heart Rate. *Acta Paediatr.* **2017**, *106*, 710–720. [CrossRef] [PubMed]
2. Rattanyu, K.; Ohkura, M.; Mizukawa, M. Emotion Monitoring from Physiological Signals for Service Robots in the Living Space. In Proceedings of the International Conference on Control, Automation and Systems, Gyeonggi-do, Republic of Korea, 27–30 October 2010; pp. 580–583. [CrossRef]
3. Cook, L.G.; Mitschow, L.C. Beyond the polygraph: Deception detection and the autonomic nervous system. *Fed. Pract.* **2019**, *36*, 316–321. [PubMed]
4. De Salles, B.F.; Simão, R.; Miranda, F.; da Silva Novaes, J.; Lemos, A.; Willardson, J.M. Rest Interval between Sets in Strength Training. *Sport. Med.* **2009**, *39*, 765–777. [CrossRef] [PubMed]
5. Cretikos, M.A.; Bellomo, R.; Hillman, K.; Chen, J.; Finfer, S.; Flabouris, A. Respiratory Rate: The Neglected Vital Sign. *Med. J. Aust.* **2008**, *188*, 657–659. [CrossRef] [PubMed]
6. Vinci, G.; Lindner, S.; Barbon, F.; Mann, S.; Hofmann, M.; Duda, A.; Weigel, R.; Koelpin, A. Six-Port Radar Sensor for Remote Respiration Rate and Heartbeat Vital-Sign Monitoring. *IEEE Trans. Microw. Theory Tech.* **2013**, *61*, 2093–2100. [CrossRef]
7. Yuill, G.M.; Saroya, D.; Yuill, S.L. A National Survey of the Provision for Patients with Latex Allergy. *Anaesthesia* **2003**, *58*, 775–777. [CrossRef]
8. Enger, E. *Concepts in Biology Ed*; Rex Bookstore, Inc.: Cagayan, Philippines, 2003.
9. Cretikos, M.; Chen, J.; Hillman, K.; Bellomo, R.; Finfer, S.; Flabouris, A. The Objective Medical Emergency Team Activation Criteria: A Case–Control Study. *Resuscitation* **2007**, *73*, 62–72. [CrossRef]
10. Hodgetts, T.J.; Kenward, G.; Vlachonikolis, I.G.; Payne, S.; Castle, N. The Identification of Risk Factors for Cardiac Arrest and Formulation of Activation Criteria to Alert a Medical Emergency Team. *Resuscitation* **2002**, *54*, 125–131. [CrossRef]
11. Goldhill, D.R.; McNarry, A.F.; Mandersloot, G.; McGinley, A. A Physiologically-Based Early Warning Score for Ward Patients: The Association between Score and Outcome. *Anaesthesia* **2005**, *60*, 547–553. [CrossRef]
12. Goldhill, D.R.; Worthington, L.; Mulcahy, A.; Tarling, M.; Sumner, A. The Patient-at-Risk Team: Identifying and Managing Seriously Ill Ward Patients. *Anaesthesia* **1999**, *54*, 853–860. [CrossRef]
13. Hourihan, F.; Bishop, G.; Hillman, K.M.; Daffurn, K. The Medical Emergency Team: A New Strategy to Identify and Intervene in High-Risk Patients. *Clin. Intensive Care* **1995**, *6*, 269–272. [CrossRef]
14. Jerath, R.; Beveridge, C. Respiratory Rhythm, Autonomic Modulation, and the Spectrum of Emotions: The Future of Emotion Recognition and Modulation. *Front. Psychol.* **2020**, *11*, 1980. [CrossRef]
15. Philippot, P.; Chapelle, G.; Blairy, S. Respiratory Feedback in the Generation of Emotion. *Cogn. Emot.* **2002**, *16*, 605–627. [CrossRef]
16. Butler, E.A.; Wilhelm, F.H.; Gross, J.J. Respiratory Sinus Arrhythmia, Emotion, and Emotion Regulation during Social Interaction: RSA and Emotion Regulation. *Psychophysiology* **2006**, *43*, 612–622. [CrossRef]
17. Sowho, M.; Amatoury, J.; Kirkness, J.P.; Patil, S.P. Sleep and Respiratory Physiology in Adults. *Clin. Chest Med.* **2014**, *35*, 469–481. [CrossRef]
18. Bianchi, W.; Dugas, A.F.; Hsieh, Y.-H.; Saheed, M.; Hill, P.; Lindauer, C.; Terzis, A.; Rothman, R.E. Revitalizing a Vital Sign: Improving Detection of Tachypnea at Primary Triage. *Ann. Emerg. Med.* **2013**, *61*, 37–43. [CrossRef]
19. Flenady, T.; Dwyer, T.; Applegarth, J. Accurate Respiratory Rates Count: So Should You! Australas. *Emerg. Nurs. J.* **2017**, *20*, 45–47. [CrossRef]
20. Młyńczak, M.; Cybulski, G. Impedance Pneumography: Is It Possible? In Proceedings of the Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments, Wilga, Poland, 28 May–3 June 2012; Romaniuk, R.S., Ed.; 2012.
21. Gupta, A.K. *Respiration Rate Measurement Based on Impedance Pneumography*; Data Acquisition Products, Application Report, SBAA181; Texas Instruments: Dallas, TX, USA, 2011.
22. Seppä, V.-P.; Pelkonen, A.S.; Kotaniemi-Syrjänen, A.; Viik, J.; Mäkelä, M.J.; Malmberg, L.P. Tidal Flow Variability Measured by Impedance Pneumography Relates to Childhood Asthma Risk. *Eur. Respir. J.* **2016**, *47*, 1687–1696. [CrossRef]
23. Jarchi, D.; Rodgers, S.J.; Tarassenko, L.; Clifton, D.A. Accelerometry-Based Estimation of Respiratory Rate for Post-Intensive Care Patient Monitoring. *IEEE Sens. J.* **2018**, *18*, 4981–4989. [CrossRef]
24. Mahbub, I.; Pullano, S.A.; Wang, H.; Islam, S.K.; Fiorillo, A.S.; To, G.; Mahfouz, M.R. A Low-Power Wireless Piezoelectric Sensor-Based Respiration Monitoring System Realized in CMOS Process. *IEEE Sens. J.* **2017**, *17*, 1858–1864. [CrossRef]
25. Hermann, S.; Lombardo, L.; Campobello, G.; Burke, M.; Donato, N. A Ballistocardiogram Acquisition System for Respiration and Heart Rate Monitoring. In Proceedings of the 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Houston, TX, USA, 14–17 May 2018; pp. 1–5.
26. Atalay, O.; Kennon, W.R.; Demirok, E. Weft-Knitted Strain Sensor for Monitoring Respiratory Rate and Its Electro-Mechanical Modeling. *IEEE Sens. J.* **2015**, *15*, 110–122. [CrossRef]
27. Norman, R.G.; Ahmed, M.M.; Walsleben, J.A.; Rapoport, D.M. Detection of respiratory events during NPSG: Nasal cannula/pressure sensor versus thermistor. *Sleep* **1997**, *20*, 1175–1184. [PubMed]
28. Mogera, U.; Sagade, A.A.; George, S.J.; Kulkarni, G.U. Ultrafast Response Humidity Sensor Using Supramolecular Nanofibre and Its Application in Monitoring Breath Humidity and Flow. *Sci. Rep.* **2014**, *4*, 4103. [CrossRef] [PubMed]
29. Sanders, A.B. Capnometry in Emergency Medicine. *Ann. Emerg. Med.* **1989**, *18*, 1287–1290. [CrossRef]
30. Liu, H.; Allen, J.; Zheng, D.; Chen, F. Recent Development of Respiratory Rate Measurement Technologies. *Physiol. Meas.* **2019**, *40*, 07TR01. [CrossRef]

31. Addison, P.S.; Watson, J.N.; Mestek, M.L.; Mecca, R.S. Developing an Algorithm for Pulse Oximetry Derived Respiratory Rate (RR_{oxi}): A Healthy Volunteer Study. *Int. J. Clin. Monit. Comput.* **2012**, *26*, 45–51. [CrossRef]
32. Yasuma, F.; Hayano, J.I. Respiratory sinus arrhythmia: Why does the heartbeat synchronize with respiratory rhythm? *Chest* **2004**, *125*, 683–690. [CrossRef]
33. Ugnell, H.; Öberg, P.Å. The time-variable photoplethysmographic signal; dependence of the heart synchronous signal on wavelength and sample volume. *Med. Eng. Phys.* **1995**, *17*, 571–578. [CrossRef]
34. De Haan, G.; van Leest, A. Improved Motion Robustness of Remote-PPG by Using the Blood Volume Pulse Signature. *Physiol. Meas.* **2014**, *35*, 1913–1926. [CrossRef]
35. Verkruysse, W.; Svaasand, L.O.; Nelson, J.S. Remote Plethysmographic Imaging Using Ambient Light. *Opt. Express* **2008**, *16*, 21434–21445. [CrossRef]
36. Meredith, D.J.; Clifton, D.; Charlton, P.; Brooks, J.; Pugh, C.W.; Tarassenko, L. Photoplethysmographic Derivation of Respiratory Rate: A Review of Relevant Physiology. *J. Med. Eng. Technol.* **2012**, *36*, 1–7. [CrossRef]
37. Karlen, W.; Raman, S.; Ansermino, J.M.; Dumont, G.A. Multiparameter Respiratory Rate Estimation from the Photoplethysmogram. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 1946–1953. [CrossRef]
38. Comon, P. Independent component analysis, a new concept? *Signal Process.* **1994**, *36*, 287–314. [CrossRef]
39. Poh, M.-Z.; McDuff, D.J.; Picard, R.W. Non-Contact, Automated Cardiac Pulse Measurements Using Video Imaging and Blind Source Separation. *Opt. Express* **2010**, *18*, 10762–10774. [CrossRef]
40. Poh, M.-Z.; McDuff, D.J.; Picard, R.W. Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 7–11. [CrossRef]
41. Felber Dietrich, D.; Schindler, C.; Schwartz, J.; Barthélémy, J.-C.; Tschopp, J.-M.; Roche, F.; von Eckardstein, A.; Brändli, O.; Leuenberger, P.; Gold, D.R.; et al. Heart Rate Variability in an Ageing Population and Its Association with Lifestyle and Cardiovascular Risk Factors: Results of the SAPALDIA Study. *Europace* **2006**, *8*, 521–529. [CrossRef]
42. Sun, Y.; Hu, S.; Azorin-Peris, V.; Greenwald, S.; Chambers, J.; Zhu, Y. Motion-Compensated Noncontact Imaging Photoplethysmography to Monitor Cardiorespiratory Status during Exercise. *J. Biomed. Opt.* **2011**, *16*, 077010. [CrossRef]
43. Wei, B.; He, X.; Zhang, C.; Wu, X. Non-Contact, Synchronous Dynamic Measurement of Respiratory Rate and Heart Rate Based on Dual Sensitive Regions. *Biomed. Eng. Online* **2017**, *16*, 17. [CrossRef]
44. Tarassenko, L.; Villarroel, M.; Guazzi, A.; Jorge, J.; Clifton, D.A.; Pugh, C. Non-Contact Video-Based Vital Sign Monitoring Using Ambient Light and Auto-Regressive Models. *Physiol. Meas.* **2014**, *35*, 807–831. [CrossRef]
45. Ghodratigohar, M. A Machine Learning-Based Non-Contact Respiratory Rate Monitoring Method Using an RGB Camera. Ph.D. Thesis, Université d'Ottawa/University of Ottawa, Ottawa, ON, Canada, 2019.
46. Sanyal, S.; Nundy, K.K. Algorithms for Monitoring Heart Rate and Respiratory Rate from the Video of a User's Face. *IEEE J. Transl. Eng. Health Med.* **2018**, *6*, 2700111. [CrossRef]
47. Van Gastel, M.; Stuijk, S.; de Haan, G. Robust Respiration Detection from Remote Photoplethysmography. *Biomed. Opt. Express* **2016**, *7*, 4941–4957. [CrossRef] [PubMed]
48. Motin, M.A.; Karmakar, C.K.; Palaniswami, M. Ensemble Empirical Mode Decomposition with Principal Component Analysis: A Novel Approach for Extracting Respiratory Rate and Heart Rate from Photoplethysmographic Signal. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 766–774. [CrossRef] [PubMed]
49. Nam, Y.; Lee, J.; Chon, K.H. Respiratory Rate Estimation from the Built-in Cameras of Smartphones and Tablets. *Ann. Biomed. Eng.* **2014**, *42*, 885–898. [CrossRef] [PubMed]
50. Karlen, W.; Garde, A.; Myers, D.; Scheffer, C.; Ansermino, J.M.; Dumont, G.A. Estimation of Respiratory Rate from Photoplethysmographic Imaging Videos Compared to Pulse Oximetry. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 1331–1338. [CrossRef] [PubMed]
51. Massaroni, C.; Nicolo, A.; Sacchetti, M.; Schena, E. Contactless Methods for Measuring Respiratory Rate: A Review. *IEEE Sens. J.* **2021**, *21*, 12821–12839. [CrossRef]
52. Lienhart, R.; Maydt, J. An Extended Set of Haar-like Features for Rapid Object Detection. In Proceedings of the International Conference on Image Processing, Rochester, NY, USA, 22–25 September 2002; Volume 1, p. I.
53. Zhao, F.; Li, M.; Qian, Y.; Tsien, J.Z. Remote Measurements of Heart and Respiration Rates for Telemedicine. *PLoS ONE* **2013**, *8*, e71384. [CrossRef]
54. Reyes, B.A.; Reljin, N.; Kong, Y.; Nam, Y.; Chon, K.H. Tidal Volume and Instantaneous Respiration Rate Estimation Using a Volumetric Surrogate Signal Acquired via a Smartphone Camera. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 764–777. [CrossRef]
55. Reyes, B.A.; Reljin, N.; Kong, Y.; Nam, Y.; Ha, S.; Chon, K.H. Towards the Development of a Mobile Phonopneumogram: Automatic Breath-Phase Classification Using Smartphones. *Ann. Biomed. Eng.* **2016**, *44*, 2746–2759. [CrossRef]
56. Prathosh, A.P.; Praveena, P.; Mestha, L.K.; Bharadwaj, S. Estimation of Respiratory Pattern from Video Using Selective Ensemble Aggregation. *IEEE Trans. Signal Process.* **2017**, *65*, 2902–2916. [CrossRef]
57. Massaroni, C.; Lo Presti, D.; Formica, D.; Silvestri, S.; Schena, E. Non-Contact Monitoring of Breathing Pattern and Respiratory Rate via RGB Signal Measurement. *Sensors* **2019**, *19*, 2758. [CrossRef]
58. Wu, H.-Y.; Rubinstein, M.; Shih, E.; Guttag, J.; Durand, F.; Freeman, W. Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Trans. Graph.* **2012**, *31*, 1–8. [CrossRef]
59. Al-Naji, A.; Chahl, J. Remote Respiratory Monitoring System Based on Developing Motion Magnification Technique. *Biomed. Signal Process. Control* **2016**, *29*, 1–10. [CrossRef]

60. Antognoli, L.; Marchionni, P.; Nobile, S.; Carnielli, V.P.; Scalise, L. Assessment of Cardio-Respiratory Rates by Non-Invasive Measurement Methods in Hospitalized Preterm Neonates. In Proceedings of the 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Rome, Italy, 11–13 June 2018; pp. 1–5.
61. Horn, B.K.P.; Schunck, B.G. Determining Optical Flow. *Artif. Intell.* **1981**, *17*, 185–203. [CrossRef]
62. Lin, K.-Y.; Chen, D.-Y.; Tsai, W.-J. Image-Based Motion-Tolerant Remote Respiratory Rate Evaluation. *IEEE Sens. J.* **2016**, *16*, 3263–3271. [CrossRef]
63. Tran, Q.-V.; Su, S.-F.; Nguyen, V.-T. Pyramidal Lucas—Kanade-Based Noncontact Breath Motion Detection. *IEEE Trans. Syst. Man Cybern. Syst.* **2020**, *50*, 2659–2670. [CrossRef]
64. Janssen, R.; Wang, W.; Moço, A.; de Haan, G. Video-Based Respiration Monitoring with Automatic Region of Interest Detection. *Physiol. Meas.* **2016**, *37*, 100–114. [CrossRef]
65. Massaroni, C.; Schena, E.; Silvestri, S.; Maji, S. Comparison of Two Methods for Estimating Respiratory Waveforms from Videos without Contact. In Proceedings of the 2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Istanbul, Turkey, 26–28 June 2019; pp. 1–6.
66. Shao, D.; Yang, Y.; Liu, C.; Tsow, F.; Yu, H.; Tao, N. Noncontact Monitoring Breathing Pattern, Exhalation Flow Rate and Pulse Transit Time. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 2760–2767. [CrossRef]
67. Jorge, J.; ViIllarroel, M.; Chaichulee, S.; McCormick, K.; Tarassenko, L. Data Fusion for Improved Camera-Based Detection of Respiration in Neonates. In Proceedings of the Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics, San Francisco, CA, USA, 29–30 January 2018.
68. Brieva, J.; Ponce, H.; Moya-Albor, E. A Contactless Respiratory Rate Estimation Method Using a Hermite Magnification Technique and Convolutional Neural Networks. *Appl. Sci.* **2020**, *10*, 607. [CrossRef]
69. Wiede, C.; Richter, J.; Manuel, M.; Hirtz, G. Remote Respiration Rate Determination in Video Data-Vital Parameter Extraction Based on Optical Flow and Principal Component Analysis. In Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Porto, Portugal, 27 February–1 March 2017.
70. Ganfure, G.O. Using Video Stream for Continuous Monitoring of Breathing Rate for General Setting. *Signal Image Video Process.* **2019**, *13*, 1395–1403. [CrossRef]
71. Mateu-Mateus, M.; Guede-Fernández, F.; Ferrer-Mileo, V.; García-González, M.A.; Ramos-Castro, J.; Fernández-Chimeno, M. Comparison of video-based methods for respiration rhythm measurement. *Biomed. Signal Process. Control.* **2019**, *51*, 138–147. [CrossRef]
72. Schrumpf, F.; Monch, C.; Bausch, G.; Fuchs, M. Exploiting Weak Head Movements for Camera-Based Respiration Detection. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; Volume 2019, pp. 6059–6062. [CrossRef]
73. Viola, P.; Jones, M. Rapid Object Detection Using a Boosted Cascade of Simple Features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, Kauai, HI, USA, 8–14 December 2001.
74. Kreiss, S.; Bertoni, L.; Alahi, A. PifPaf: Composite Fields for Human Pose Estimation. *arXiv* **2019**. [CrossRef]
75. Bazarevsky, V.; Grishchenko, I.; Raveendran, K.; Zhu, T.; Zhang, F.; Grundmann, M. BlazePose: On-Device Real-Time Body Pose Tracking. *arXiv* **2020**. [CrossRef]
76. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. *arXiv* **2013**. [CrossRef]
77. Hubel, D.H.; Wiesel, T.N. Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *J. Physiol.* **1962**, *160*, 106–154. [CrossRef]
78. Luo, W.; Li, Y.; Urtasun, R.; Zemel, R. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. *arXiv* **2017**. [CrossRef]
79. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *arXiv* **2014**, arXiv:1412.7062.
80. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *arXiv* **2016**, arXiv:1606.00915. [CrossRef]
81. Lawal, M.O. Tomato Detection Based on Modified YOLOv3 Framework. *Sci. Rep.* **2021**, *11*, 1447. [CrossRef]
82. Chandio, A.; Gui, G.; Kumar, T.; Ullah, I.; Ranjbarzadeh, R.; Roy, A.M.; Hussain, A.; Shen, Y. Precise Single-Stage Detector. *arXiv* **2022**, arXiv:2210.04252.