

# Joint Use of a Low Thermal Resolution Thermal Camera and an RGB Camera for Respiration Measurement

Zaeed Khan<sup>1</sup>, Matias Rusanen<sup>2</sup>, Miika Arvonen<sup>3</sup>, Timo Leppänen<sup>4</sup>, and Simo Särkkä<sup>5</sup>, *Senior Member, IEEE*

**Abstract**—Recently, joint use of thermal camera and RGB camera has emerged as a potential solution for respiration measurement. Bulky and large thermal cameras with high thermal resolution have been mostly used in thermal-imaging-based respiration measurement studies. We aim to present a method for respiration measurement using an RGB camera and a low thermal resolution thermal camera. Thermal cameras with low thermal resolution are small in size and easily portable, allowing flexible and a simple implementation of the measurement system. First, the cameras were calibrated and synchronized using calibration objects and sound analysis tools, respectively. Second, a cross-spectrum mapping method was developed which was followed by the extraction of the respiration signal. Third, the extracted and reference respiration signals were aligned. In the last step of the proposed method, a signal processing framework was applied to the estimated and reference respiration signals. The performance of the cross-spectrum mapping method was visually evaluated. The mapping method performed well when a participant had little to no head movements; however, the performance decreased when head movements were present. Our analysis showed that there is a relatively good agreement between the proposed and the reference respiration measurement methods. The number of breathing cycles from the estimated respiration signal was computed with a decent accuracy. However, the accuracy of measuring breathing rate is low compared to the previous studies. The estimated breathing patterns were measured with better accuracy than in the previous work. This study demonstrates that respiration measurement can be done by the joint use of an RGB camera and thermal camera that has a low thermal resolution, with a decent level of accuracy.

**Index Terms**—Breathing measurement, camera calibration, camera synchronization, noncontact method, RGB imaging, thermal imaging.

## I. INTRODUCTION

RESPIRATION signal carries valuable information about health conditions [1]. For example, an increase in respiratory rate (RR) may indicate heart or lung diseases, whereas a decrease in RR is generally linked to hypothermia or diseases that are related to the central nervous system [2]. Also, depth of breathing has been associated with some illnesses [3], and sleep-disordered breathing is highly prevalent in the general population [4]. Besides the medical applications, respiration measurements can be conducted during a physical exercise to assess physical effort and exercise tolerance [5]. Since respiration is an important indicator of health conditions, respiration measurement methods have become an important research topic in the domains of health and sports technology.

Respiration measurement methods are typically divided into two categories: contact- and noncontact-based measurement methods [6]. Traditional measurements require physical contact with individuals which may cause discomfort and stress [7]. This, in turn, affects negatively the measurement accuracy and quality. Therefore, noncontact measurement methods have been proposed to improve user-friendliness and measurement accuracy. Microphones, radar sensors, RGB cameras, and thermal cameras can be employed to obtain respiration signals in a noncontact way [8]. In this article, thermal-imaging-based respiration measurement is focused on considering its advantages, for example, being a passive method and independent of environmental illumination [6]. The thermal-imaging-based respiration measurement utilizes the temperature variation of nostrils which correlates with inhalation and exhalation. During inhalation, air enters into nostrils cooling down the nostril temperature. On the other hand, air exits the nostrils during exhalation, and the nostril temperature rises. Despite the advantages, thermal-imaging-based respiration measurement has also limitations that affect the accuracy of the measurement. Thermal images provide few geometric and textural facial details [7] which is a limiting factor for designing fast and reliable nostril detection and tracking algorithms.

Recently, a few papers [2], [7], [9], [10], [11], [12], [13] have proposed a hybrid method for the respiration

Manuscript received 3 July 2023; revised 15 September 2023; accepted 26 October 2023. Date of publication 28 November 2023; date of current version 29 December 2023. This work was supported in part by the Academy of Finland and the Finnish Cultural Foundation through the Kainuu Regional Fund and in part by the Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding under Grant 504807. The Associate Editor coordinating the review process was Dr. Lihui Peng. (Corresponding author: Zaeed Khan.)

Zaeed Khan and Simo Särkkä are with the Department of Electrical Engineering and Automation, Aalto University, 02150 Espoo, Finland (e-mail: zaeed.khan@aalto.fi; simo.sarkka@aalto.fi).

Matias Rusanen is with the Department of Technical Physics, University of Eastern Finland, 70210 Kuopio, Finland, and also with the Diagnostic Imaging Center, Kuopio University Hospital, 70200 Kuopio, Finland (e-mail: matias.rusanen@uef.fi).

Miika Arvonen is with the Department of Pediatrics, Kuopio University Hospital, 70200 Kuopio, Finland (e-mail: miika.arvonen@kuh.fi).

Timo Leppänen is with the Department of Technical Physics, University of Eastern Finland, 70210 Kuopio, Finland, also with the Diagnostic Imaging Center, Kuopio University Hospital, 70200 Kuopio, Finland, and also with the School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia (e-mail: timo.leppanen@uef.fi).

Digital Object Identifier 10.1109/TIM.2023.3334345

measurement that employs both thermal camera and RGB camera. The purpose of adding an RGB camera is to allow more accurate and faster detection and tracking of nostrils and other facial features in thermal images. Because of the red (R), green (G), and blue (B) spectrum illumination, the RGB images provide richer information about the face compared to the thermal images [12]. Based on the previous studies, the respiration measurement process that utilizes the simultaneous use of RGB camera and thermal camera consists of the following steps: nostril detection in the first RGB video frame, tracking of the region of interest (ROI), cross-spectral mapping of the ROI from an RGB video frame to corresponding thermal video frame, and respiration signal extraction from thermal video frames.

In most of the previous studies, relatively large and bulky thermal cameras have been used. Large thermal cameras have high thermal resolutions, such as  $320 \times 240$  and  $640 \times 480$ , and many other features. Regardless, high thermal resolution thermal cameras are often bulky and heavy which make them less portable, and they are also more expensive which makes them less accessible. In our work, we used a low thermal resolution thermal camera, which is smaller in size, less expensive, more portable, and can be connected to a smartphone. However, due to the low thermal resolution ( $160 \times 120$ ), the clarity in the produced thermal images is worse.

The aim of this work was to develop a respiration measurement method using recordings of a low thermal resolution thermal camera and an RGB camera which would detect a number of breathing cycles in a given period as well as measure different breathing patterns that are clinically significant, such as tachypnea and Kussmaul breathing. Contributions of the proposed work were: 1) the measurement of breathing with a low thermal resolution thermal camera; 2) the development of calibration and synchronization processes between the cameras; 3) the development of calibration and synchronization processes between the proposed and reference methods; 4) the development of cross-spectrum tracking that is based on the camera calibration process and different pixel intensities of different facial features in the thermal video; 5) development of signal processing pipeline; and 6) the evaluation of the calibration process, cross-spectral mapping, alignment of signals and the performance of the proposed method in breathing rate measurement, computation of the number of breathing cycles, and breathing pattern measurement.

## II. RELATED WORK

In this section, we review breathing measurement methods using multicamera systems and cross-spectrum mapping methods used in previous RGB-thermal-imaging-based respiration measurements.

### A. Multicamera Vision-Based Respiration Measurement

Recently, multicamera systems have been utilized for respiration measurement. Deng et al. [14] and Lorato et al. [15] proposed multicamera system-based respiration measurement for sleep monitoring and infant respiration monitoring, respectively. Deng et al. employed six infrared cameras as well as

a Kinect motion sensor that has a color camera, an infrared projector, and a depth camera. The infrared cameras were used for head tracking, and the Kinect sensor was used for body posture recognition. Lorato et al. used three thermal cameras for collecting thermal videos of infants in open bed. However, the above-mentioned work does not deal with cross-spectrum mapping that is presented in a few papers. Hu et al. [7] used an MAG62 thermal camera that has a thermal resolution of  $640 \times 480$ , with an RGB camera for respiration rate and breathing pattern measurements. Chen et al. also used MAG62 thermal camera in papers [2] and [9], both of which present respiration rate measurement methods using thermal and RGB cameras simultaneously. Maurya et al. [10], [12] proposed techniques for respiration monitoring that includes respiration rate and breathing pattern measurements, by integrating RGB and thermal imaging. For the thermal imaging, Maurya et al. used an FLIR E-60 with a thermal resolution of  $320 \times 240$ . The simultaneous use of RGB and thermal cameras for respiration rate measurement is also proposed by Negishi et al. [11], [13] who used an FLIR A315 thermal camera that has also the thermal resolution of  $320 \times 240$ , for extracting the breathing signal.

### B. Cross-Spectrum Mapping in RGB Thermal Imaging

Cross-spectrum mapping is a fundamental step in an RGB-thermal-imaging-based respiration measurement, since it maps an ROI from the RGB video frame to the corresponding thermal video frame, and therefore, it is an important step in respiration measurement methods using RGB and thermal imaging simultaneously. Chen et al. [2], [9], Hu et al. [7], and Maurya et al. [10], [12] used linear coordinate cross-spectrum mapping, which is a simple mapping method. However, it required an image transformation on RGB video frames in order to resize the row and column of the RGB video frames equal to the corresponding thermal video frames which enabled the linear coordinate cross-spectrum mapping. Maurya et al. used a fine registration algorithm in addition to the image registration for refining the alignment, since the limitation of control points selected in a calibration object may lead to some misalignment. Negishi et al. [13] developed a linear coordinate cross-spectrum mapping without image transformation of RGB video frames assuming that the distance from the camera to the subject is 1 m all the time. Negishi et al. [11] also developed another cross-spectrum mapping method that estimates a homography matrix between the image coordinates in the RGB video frame and the corresponding thermal video frame.

## III. MATERIALS AND METHODS

In this section, we describe the experimental setup, protocols, camera calibration processes, camera-belt calibration process, ROI detection and tracking method, cross-spectrum mapping method, extraction of breathing signal, as well as signal processing.

### A. Experimental Setup

Respiration measurements were conducted at the Smart-Sleep Laboratory at the University of Eastern Finland. Seven

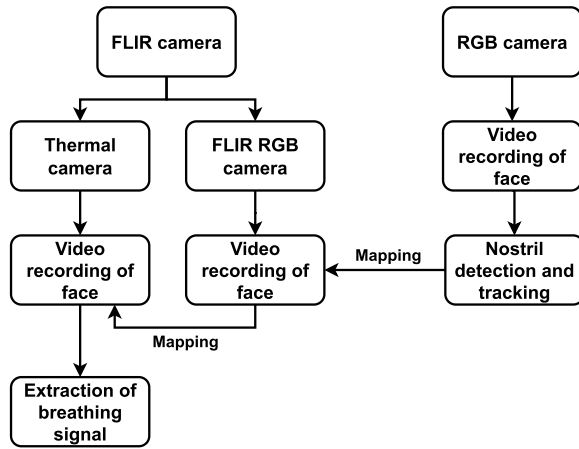


Fig. 1. In the proposed breathing measurement method, three cameras were used: an RGB camera, an FLIR RGB camera, and a thermal camera. The RGB camera was used to detect and track the ROI throughout the video recordings. The FLIR RGB video and the thermal video recordings were fused into one video which facilitated the mapping of the ROI from the RGB video frame to the respective thermal video frame via the FLIR RGB video frame. The breathing signal was extracted from the thermal video.

healthy volunteers participated in the experiments. All participants were males with the age range of 24–46. In data analysis, we had to exclude the measurement data of one participant because the participant shifted himself during the recording which invalidated the calibration. Therefore, there are six participants. The framework of the respiration measurement method that was employed in this study is shown in Fig. 1. The experimental protocol was given a favorable statement by The Research Ethics Committee of the Northern Savo Hospital District (permit number FIMEA/2020/007208), and participants gave written informed consent before the experiments.

In the experimental setup, the following tools and devices were used: two tripods, two smartphones, one FLIR ONE Pro (Teledyne FLIR LLC, Wilsonville, OR, USA) camera, two respiratory inductance plethysmography (RIP) belts, and Nox A1 recording device (Nox Medical, Reykjavik, Iceland) as well as two calibration objects, and a laptop. In this work, one of the smartphones was used for RGB video recording, and the other one was used for thermal video recording. The reason behind the use of two smartphones was not having a customized mobile application that would enable simultaneous use of the smartphone camera and the thermal camera attached to it. The FLIR One Pro camera device (see Fig. 2) consists of two cameras: an RGB camera and a thermal camera with a thermal resolution of  $160 \times 120$ . Although the FLIR One Pro has an RGB camera, we did not use it for nostril detection and tracking due to the low resolution of the camera. The low resolution would make the camera prone to failure in detecting the facial landmarks and ROI as well as keeping track of the ROI. We also could not exclude the use of FLIR’s RGB camera because calibration between the thermal camera and the smartphone camera would require the participants to hold a warm object which is not user-friendly. Also, the thermal camera has so low resolution that calibration points on calibration object are not clearly visible on the thermal video. To avoid any confusion in this article, the smartphone



Fig. 2. FLIR One Pro camera device consists of two cameras: an RGB camera (top) and a thermal camera (bottom).



Fig. 3. Cross-spectral mapping of ROI is a two-step process: in the first step, the ROI is mapped from RGB video to FLIR RGB video, and in the second step, the ROI is again mapped from FLIR RGB video to thermal video.

camera, the RGB camera of the FLIR One Pro, and the thermal camera of the FLIR One Pro will be referred to as RGB camera, FLIR RGB camera, and thermal camera, respectively. During the measurements, all three aforementioned cameras were used. Since both cameras of the FLIR camera device were in use, multispectral dynamic imaging (MSX) was enabled which added visible light details (e.g., edges) to thermal images or videos [16] that is, in MSX, the FLIR RGB images and the thermal images were fused into one image [see Section III-C]. The calibration objects were A4-size papers, one with grids marked with black marker and one with square holes. The calibration object with grids was used for RGB–FLIR RGB camera calibration, whereas the calibration with a square hole was used for FLIR RGB–thermal camera calibration.

The smartphones were placed on the tripods, and the FLIR One Pro camera device was connected to one of the smartphones. The smartphone, to which the FLIR camera was not connected, was used for nostril detection and movement tracking, while the thermal camera was used for extracting breathing signals. The FLIR RGB camera was used for performing calibration between the RGB camera and the thermal camera in order to map ROI from the RGB image sequence to the respective thermal image sequence. The flowchart of the cross-spectral mapping is presented in Fig. 3. The camera calibration process consists of two steps: RGB-FLIR RGB camera calibration and FLIR RGB-thermal camera calibration. The RGB-FLIR RGB camera calibration was performed during each measurement while the FLIR RGB-thermal camera calibration was only performed once outside the measurements. The main reason for the FLIR RGB-thermal camera calibration is that the cameras of the FLIR device were not aligned. The RIP belts were used for the reference measurements. One respiration belt was put around the thorax, while the other was put around the abdomen. Calibration objects were used for obtaining a mapping relation between the cameras. The tripods were set next to each other and participants sat in front of the tripods.

## B. Protocols

The respiration experiment consists of four protocols. Participants performed all four sets in one run (i.e., video recordings and respiration belt recordings were not stopped between the protocols). During each protocol, participants could only breathe through the nose. Before starting a new experiment, necessary distance measurements were done for the subject. The following distances were measured to adjust the position of subjects in relation to the cameras: smartphone camera to subject's nose (O-S), FLIR camera to subject's nose (T-S), and seat to subject's nose (S-S). The ranges of distances were 72–78, 72–77, and 76–83 cm for O-S, T-S, and S-S distances, respectively. At the beginning of each experiment, an RGB camera–FLIR RGB camera calibration and an RGB–respiration belt calibration were performed. The experimental setup is illustrated in Fig. 4.

The experimental protocols were as follows.

- 1) In the first protocol, participants breathed normally for 1 min. Participants tried to avoid head movements while breathing.
- 2) Participants again breathed normally, but this time, they could move their heads freely. The duration of the protocol was 1 min.
- 3) Participants were asked to put their tongues toward the palate and to perform deep inhalations. Thus, air cannot move properly but nose wings could move. Participants were asked to perform around ten cycles of such breathing.
- 4) In the fourth protocol, participants breathed by following breathing patterns on a video. The 1-min-long video shows breathing patterns in the following order: apnea (0–3 s), eupnea (3–17 s), tachypnea (17–24 s), apnea (24–28 s), eupnea (28–41 s), Kussmaul (41–51 s), and apnea (51–60 s). A custom Breath Dictator program that was developed for simulating different breathing patterns was used to generate the particular breathing pattern sequence. Participants were asked to inhale when the marker on the breathing signal was going upward along the curve and exhale when the marker was going down along the curve.

## C. Smartphone-FLIR Camera Calibration

The purpose of the calibration between the RGB camera and FLIR RGB camera was to map correctly the nostril region from RGB image sequence to the corresponding thermal image sequence. Before starting the video recordings, participants held the calibration object below their chin and above the chest, and they were asked to say “calibration begins” while holding the calibration object after the cameras started recording.

Using sound analysis tools (Python's PyDub library [17]), we extracted the time instants when the participant started to speak from both RGB video as well as thermal video. Let the time instant for RGB video be denoted with  $t_{\text{RGB}}$ , and for thermal video, it is denoted with  $t_{\text{therm}}$ . The delay between videos was computed by simple subtraction operation

$$t_{\text{delta}} = t_{\text{RGB}} - t_{\text{therm}}. \quad (1)$$

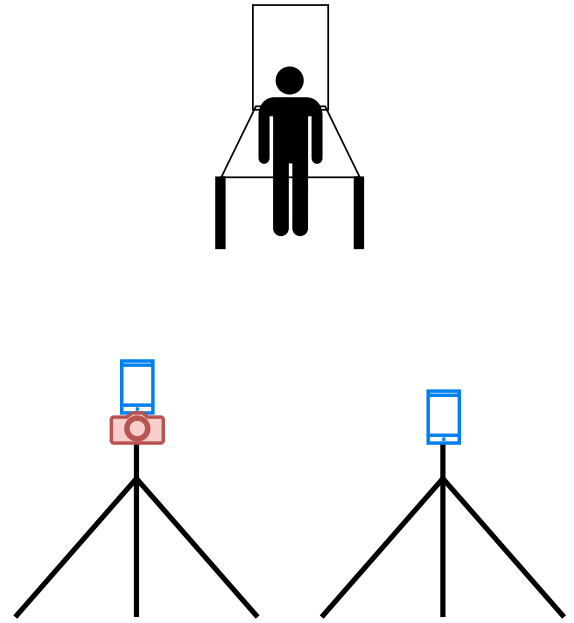


Fig. 4. Experimental setup for the breath detection experiment is illustrated in this figure. Two tripods were used. The tripod on the left held a smartphone with the thermal camera, and the tripod on the right held a smartphone only. A subject sat in front of the cameras.

Next, frame rates of both videos were extracted (using Python's OpenCV library [18]), and they were used to obtain an RGB video frame and an FLIR video frame (fusion of FLIR RGB and thermal video data) corresponding to  $t_{\text{RGB}}$  and  $t_{\text{therm}}$ , respectively, by following equation:

$$N = \text{round}(ft) \quad (2)$$

where  $f$  is the frame rate,  $t$  is the time instant, and  $N$  is a positive integer that refers to the index of the video frame (e.g.,  $N = 1$  means the first video frame). Hence, we could assume that the scenes in the obtained video frames are from the same moment, and they contain the calibration object. The delay information in (1) was used to retrieve the FLIR video frame corresponding to the color video frame. For example, if we have  $n$ th RGB video frame, the index of the corresponding FLIR video frame would be

$$m = \left( \frac{n}{f_{\text{RGB}}} - t_{\text{delta}} \right) f_{\text{therm}} \quad (3)$$

where  $f_{\text{RGB}}$  and  $f_{\text{therm}}$  are the frame rates of RGB and FLIR videos, respectively.

Calibration points were manually selected first from the RGB frame [see Fig. 5(a)]. Then, the corresponding calibration points were selected from the FLIR video frame in the same order as were in the RGB frame [see Fig. 5(b)]. The calibration points of one frame were stored as homogeneous coordinate points  $(x, y, 1)$  in a  $3 \times n$  matrix, where  $n$  is the number of selected calibration points. Let us denote the RGB calibration point matrix and FLIR calibration point matrix with  $\mathbf{X}_{\text{CRGB}}$  and  $\mathbf{X}_{\text{Ctherm}}$ , respectively. The relation between the two matrices can be described mathematically in the following way:

$$\mathbf{X}_{\text{Ctherm}} = \mathbf{L}\mathbf{X}_{\text{CRGB}} \quad (4)$$



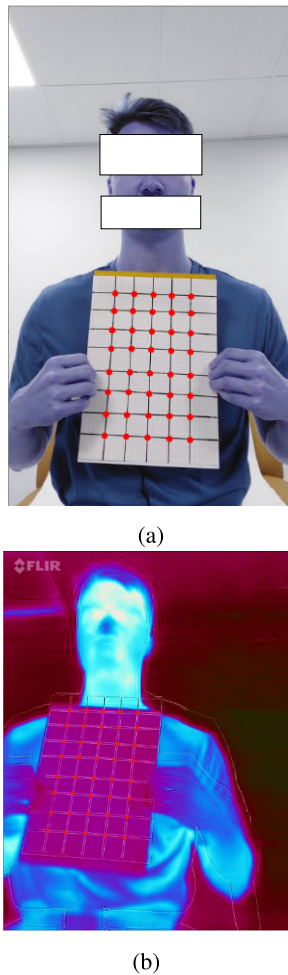


Fig. 5. Corresponding calibration points were selected manually from (a) RGB video frame and (b) corresponding FLIR video frame in the same order.

where  $\mathbf{L}$  is a  $3 \times 3$  mapping matrix that maps points from the RGB video frame to the respective FLIR video frame. For each subject, the mapping matrix was obtained by multiplying the FLIR calibration point matrix with the pseudoinverse of the RGB calibration point matrix

$$\mathbf{L} = \mathbf{X}_{\text{Ctherm}} \mathbf{X}_{\text{CRGB}}^+ \quad (5)$$

#### D. FLIR Camera Calibration

The calibration between FLIR internal cameras (FLIR RGB and thermal cameras) was performed in a similar way as the smartphone-FLIR camera calibration except different calibration object was used and the calibration was only performed once after all the measurements. The calibration object was an A4-size paper with square holes. A warm object (a laptop was used as a warm object in this work) was placed behind the calibration object so that the holes can be seen more clearly from the thermal image. A photograph of the calibration object was taken with both a thermal camera and FLIR RGB camera. From both images, respective corner points of square holes were selected in corresponding order. After having two sets of calibration points, the mapping matrix between FLIR internal cameras was solved by utilizing an analogous

equation to (5)

$$\mathbf{L}_{\text{FLIR}} = \mathbf{X}_{\text{Ftherm}} \mathbf{X}_{\text{FRGB}}^+ \quad (6)$$

where  $\mathbf{L}_{\text{FLIR}}$  is a  $3 \times 3$  mapping matrix (that maps points from FLIR RGB image to the respective thermal image),  $\mathbf{X}_{\text{Ftherm}}$  is thermal calibration point matrix, and  $\mathbf{X}_{\text{FRGB}}$  is FLIR RGB calibration point matrix. After both mapping matrices ( $\mathbf{L}$  and  $\mathbf{L}_{\text{FLIR}}$ ) are solved, an ROI can be mapped from the smartphone RGB video frame to the respective thermal frame.

#### E. Camera-Belt Calibration

The smartphone-FLIR camera calibration was followed by camera-belt calibration. As mentioned earlier, RIP belts were used as a reference measurement method for breathing detection in this work. The RIP belts and the camera system started data recording at different times which raised a need for time-synchronization between the RIP belts and the camera system. The purpose of the camera-belt calibration was to align the respiration signals yielded by the RIP belts and RGB-thermal camera system.

In camera-belt calibration, participants pulled and released the upper RIP belt three times. Fig. 6(a) shows standardized breathing signals that are not aligned, obtained by the thermal video (blue) and the upper RIP belt (red). As we can see, the signal obtained by the RIP belt has three distinctive peaks (after 100 s) that represent the upper RIP belt pulling events. In the video recording, the time instants for the RIP belt pulling events were estimated visually, and Fig. 6 shows them as three black vertical lines. Utilizing the time instants of the RIP belt pulling events, the upper RIP belt signal was aligned with respect to the breathing signal extracted from the thermal video [see Fig. 6(b)].

#### F. Nostril Detection and Tracking in RGB Video

The framework of the nostril detection method used in this work is presented in Fig. 7. In the first step, a pretrained face detector that consists of a single shot-multibox detector (SSD) and ResNet-10 architecture is used for the face detection. The face detector gives coordinates of the top-left corner as well as the width and height of the bounding box that defines the face region.

The second step is to detect facial landmarks within the face region. The facial landmark detector that was used in this work is based on [19]. In this work, we detected the facial landmarks around the nostrils.

Based on the detected facial landmarks, we defined a rectangle-shaped bounding box that contains both nostrils, and it was our ROI. After the detection of the ROI, the bounding box was tracked by Python's built-in discriminative correlation filter tracker with channel and spatial reliability [20]. Since there was a possibility for a tracker to drift away, the ROI detection and the tracker were initialized every 50 frames.

#### G. Cross-Spectrum Mapping of Nostrils

In cross-spectrum nostrils mapping, the nostrils are tracked in the thermal video by mapping the ROI from the given

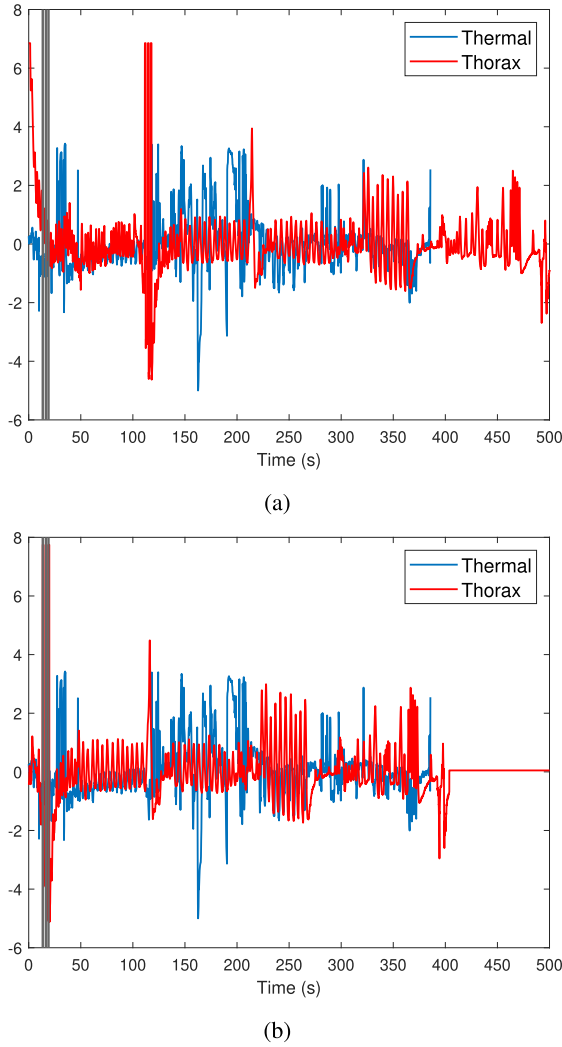


Fig. 6. (a) Since the upper RIP belt and the thermal camera started recording at different times, breathing signals obtained by the RIP belt and thermal camera were not aligned. In order to align the signals, participants pulled the upper RIP belt which was observable in both the RIP belt signal and in the thermal video. (b) Upper RIP belt signal was aligned with respect to the breathing signal obtained from the thermal video. The black vertical lines refer to the time instants that were visually estimated from the thermal video for the upper RIP belt pulling events.



Fig. 7. Method for the detection of nostrils consists of three steps: 1) detection of face; 2) detection of facial landmarks around the nostrils; and 3) detection of ROI based on the facial landmarks.

RGB video frame to the corresponding thermal video frame. Basically, the coordinates of corner points of the ROI were mapped to thermal video frames throughout the video playback, while the ROI was tracked in the RGB video. Suppose that the homogeneous coordinates of corners of the ROI in a given RGB video frame are denoted as  $\mathbf{P}$  ( $3 \times 4$  matrix). The respective coordinates of the corners in the corresponding FLIR RGB video frame would be

$$\mathbf{P}_{\text{temp}} = \mathbf{L}\mathbf{P} \quad (7)$$

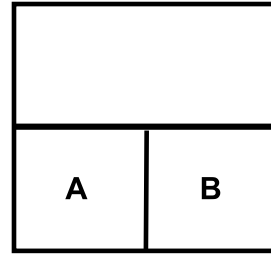


Fig. 8. In horizontal shift, the bounding box was first divided into two equal regions by a horizontal line. Second, the lower region of the bounding box was divided into two equal regions (A and B) by a vertical line. By comparing the average pixel intensity value of regions A and B, the bounding box was horizontally adjusted in the thermal video.

where  $\mathbf{L}$  is the mapping matrix obtained by (5). Then, the coordinates are further mapped to the thermal camera frame

$$\mathbf{P}_{\text{final}} = \mathbf{L}_{\text{FLIR}}\mathbf{P}_{\text{temp}} \quad (8)$$

where  $\mathbf{L}_{\text{FLIR}}$  is the mapping matrix obtained by (6). Before performing the second mapping, the given thermal video frame was changed to grayscale.

Mapping of the bounding box from the FLIR RGB camera to the thermal camera was slightly inaccurate, since the calibration between the FLIR RGB camera and thermal was only done once unlike the RGB-FLIR camera calibration. This would result in poor breathing signal quality. To overcome the mapping problem, a fine-tuning method was implemented. The implemented fine-tuning method is based on the differences between pixel intensities, and it adjusts the location of the bounding box in the given thermal video frame so that both nostrils are within the bounding box.

The fine-tuning method consists of two steps which are horizontal shift and vertical shift. The first step is the horizontal shift that shifts the bounding box in the horizontal direction so that its vertical central axis runs along the center of the nose or as close as possible. The horizontal shift is applied on grayscale thermal video frame. First, the bounding box was divided into two equal regions by a horizontal line, and the lower part of the bounding box is divided again into two equal regions that are called A and B, by a vertical line (see Fig. 8). The temperature of the nose is relatively lower than its neighboring facial parts [21]. Therefore, the pixels on the nose are darker than the pixels on the cheeks in the grayscale thermal video frame. Regions A and B are compared to each other in terms of average pixel intensity value. The darker the pixel, the lower the pixel value. If the absolute difference between the average pixel intensity values is less than a given threshold, the vertical central axis is considered to be close enough to the center line of the nose. Otherwise, the bounding box is shifted to either right or left by one pixel depending on which cell has a greater average pixel intensity value. The horizontal shifting process is performed until the absolute difference is less than the threshold. In this work, the threshold was 10.

After the horizontal shift, the bounding box might not contain the nostrils. Therefore, a vertical shift process is required to shift the bounding box in the vertical direction so that it contains the nostrils. As the horizontal shift, the

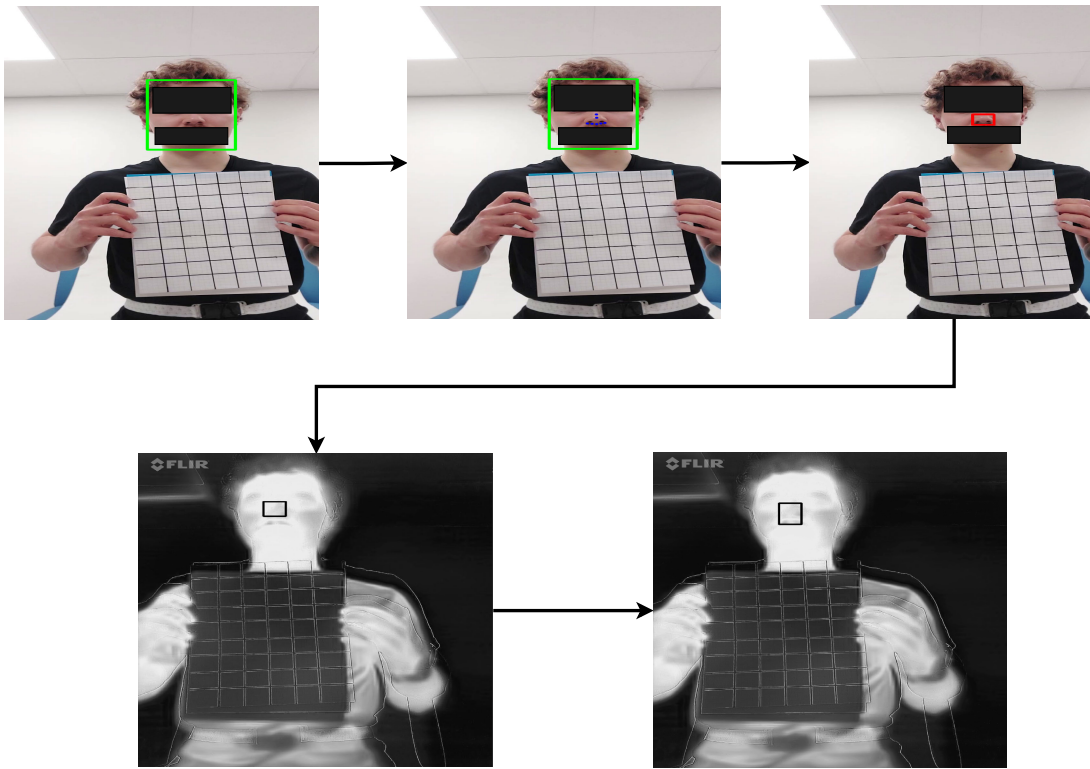


Fig. 9. Image sequence illustrates the cross-spectrum mapping of the ROI starting from the top-left image. Top-left: face of the participant is first detected from the given RGB video frame. Top-middle: few facial landmarks around the nostrils are detected. Top-right: based on the detected facial landmarks, the ROI is defined by a bounding box that is tracked in the RGB video. Bottom-left: bounding box is mapped from the RGB video frame to the corresponding thermal video frame (grayscale format). Bottom-right: location of the mapped bounding box is adjusted by applying the fine-tuning method.

vertical shift is also based on the differences between the pixel intensities. However, the vertical process is performed on a binary image that has only two types of pixels: black (0) and white (255). Thus, the given grayscale thermal video frame is converted to a binary frame. Our threshold value for the binary conversion was 225, which was determined heuristically. In the binary frame, we kept shifting the bounding box until one of the following conditions was met.

- 1) The last row of the bounding box has more white pixels than black pixels.
- 2) The bounding box is shifted ten times.

If there was no vertical shift in the process, the bounding box was shifted by ten pixels vertically along the negative  $y$ -axis. After the shifting process, the height of the bounding box was increased by 30 pixels in order to ensure that the nostrils are inside the bounding box. The fine-tuning process was applied on every thermal video frame that corresponded to every 20th frame of the RGB video frame. The detection and cross-spectrum mapping of the ROI are illustrated in Fig. 9.

#### H. Extraction of Breathing Signal

In this work, we extracted breathing signals from thermal video in grayscale format frame by frame, because each video frame corresponds to a specific timestamp. A sample of breathing signal was obtained by computing an average pixel intensity value of ROI confined by a bounding box. This method has also been applied in previous works, such as [7], [9], [10], and [12]. The following equation shows how

a sample of breathing signal was obtained from a given thermal video frame:

$$S = \frac{1}{N} \sum_{x,y \in \text{ROI}} s(x, y) \quad (9)$$

where  $s(i, j)$  is the pixel intensity value at pixel  $(x, y)$ .

#### I. Signal Processing

The breathing signals from thermal videos were filtered and analyzed in order to remove noises and in order to extract meaningful information. In this work, the signal filtering and processing were performed in MATLAB (R2022b). The breathing signal of each participant was divided into protocol parts, and each part was filtered and analyzed one at a time. Our signal processing framework consists of four steps:

- 1) signal standardization;
- 2) median filtering;
- 3) interpolation;
- 4) Butterworth filtering.

Steps 2–4 were applied to the standardized breathing signal, which has a mean of 0 and a standard deviation of 1. The formula for standardization is

$$z = \frac{x - \mu}{\sigma} \quad (10)$$

where  $z$  is the standardized value,  $x$  is the original value,  $\mu$  is the mean of the signal, and  $\sigma$  is the standard deviation of

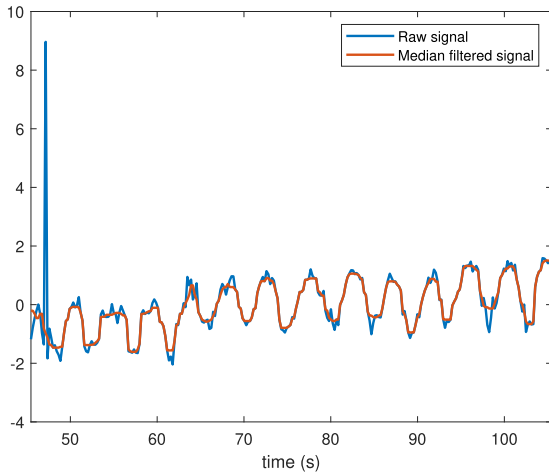


Fig. 10. Noise in the raw breathing signal (blue) is seen as a high peak. After applying a median filter on the raw signal, the noise peak disappears, and the signal becomes clearer (orange).

the signal. The standardized value is unitless. Steps 2–4 were applied to the standardized breathing signal. The breathing signals might contain noises with extremely large magnitudes. Such noises were suppressed by applying a median filter, which is a well-known filtering method for suppressing impulsive noises [22]. In our work, the length of the median filter was 7. In Fig. 10, we can see that there is a noise peak at the beginning of the raw breathing signal. After applying a median filter to the signal, the noise peak disappears, and the signal is more clear.

The median filtering was followed by interpolation. In this work, we compared the breathing signal obtained by the proposed method to the reference signal obtained by the upper RIP belt. Therefore, both signals should have the same sampling rate. However, the frame rate of the thermal camera was much lower than the sampling rate of the RIP belts which raised the need for the interpolation. In our work, we interpolated the breathing signal obtained by the proposed method in reference to the upper RIP belt signal.

After the interpolation, the quality of both breathing signals (thermal video and reference) was enhanced by a Butterworth filter. We designed the Butterworth filter in MATLAB. We used the following parameters for our filter design:

- 1) filter degree: 4;
- 2) normalized cutoff frequency:  $2f_c/f_s$ ;
- 3) filter type: high pass.

The normalized cutoff frequency is expressed with  $f_c$  (cutoff frequency) and  $f_s$  (sampling frequency). In this work,  $f_c = 0.1$  Hz.

#### J. Breathing Cycles and Breathing Rate

The number of breathing cycles was computed only from the estimated breathing signals as well as from the corresponding reference signals that were obtained during the first and the third measurement protocols. The number of breathing cycles was calculated with the following formula:

$$N_c = N_p - 1 \quad (11)$$

TABLE I  
 $R^2$  VALUES IN THE TABLE MEASURE THE LINEARITY IN THE RELATIONSHIP BETWEEN THE BREATHING SIGNAL SAMPLES OBTAINED BY THE REFERENCE METHOD AND THE PROPOSED METHOD DURING PROTOCOLS 1, 3, AND 4

Subject	Protocol 1	Protocol 3	Protocol 4
1	0.2988	0.1938	0.0291
2	0.0851	0.4630	0.1239
3	0.4496	0.2877	0.0186
4	0.2361	0.0048	0.0914
5	0.6784	0.0892	0.0248
6	0.5180	0.0500	0.0132

where  $N_p$  is the number of breathing signal peaks. The breathing rate, which is defined as the number of breathing cycles in 1 min, is computed by the following formula:

$$\text{br} = 60 \frac{N_c}{\Delta T} \quad (12)$$

where  $\Delta T$  is the time interval from which the number of breathing cycles is computed.

## IV. RESULTS

In this section, the results of this work are presented by demonstrating the performance of the proposed method in comparison to the reference method.

#### A. Evaluation of Cross-Spectral Tracking

The performance of the cross-spectrum mapping was evaluated visually in this work. When a subject did not move at all or had little head movement, the cross-spectrum mapping performed well. However, when a subject turned his head around or moved his hand toward the face, the cross-spectral mapping could not properly locate the ROI in the thermal video. Therefore, the cross-spectrum mapping performed well in protocols 1 (normal breathing without head movements), 3 (deep inhaling while having the tongue toward the palate), and 4 (simulating a breathing pattern sequence), whereas in protocol 2 (normal breathing while moving the head freely), the performance was poor.

#### B. Evaluation of Signal Alignment

For each individual breathing measurement, the performance of camera-belt calibration, which was used for aligning the estimated breathing signal with respect to the reference signal, was evaluated by linear regression method and cross correlation. Based on Table I, coefficients of determination are relatively low ( $\leq 0.66$ ) which indicates that there is a weak linear correlation between the estimated and reference breathing signal sample points. Regression lines with the best coefficients of determination from each protocol are illustrated in Fig. 11, from which it can be seen that the relations between estimated breathing signal samples and its corresponding reference signal samples are inverse (slope of the regression line is negative) in top two plots, whereas, in the bottom plot, the relationship is directly proportional (slope of the regression line is positive).



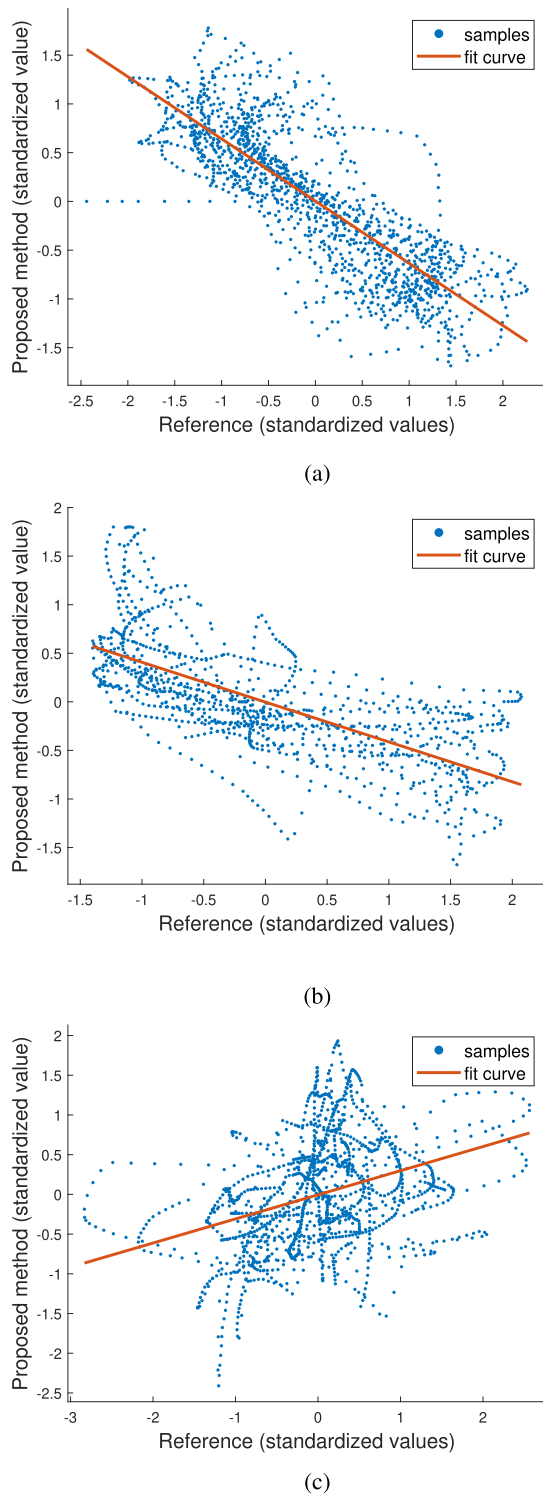


Fig. 11. Regression lines with best coefficients of determination from measurement protocols 1, 3, and 4 are illustrated along with the scatter plots of breathing signal samples (blue dots). (a) Protocol 1:  $R^2 = 0.6784$ . (b) Protocol 3:  $R^2 = 0.4630$ . (c) Protocol 4:  $R^2 = 0.1239$ .

Time lags corresponding to the minimum cross-correlations between the estimated and the corresponding breathing signals are presented in Table II. For protocol 1 of subject 2, a relatively large time lag (18.2 s) was observed compared to other time lags. The lowest time lag, which is 0.12 s, is observed for protocol 1 of subject 5.

TABLE II

TIME LAGS (IN SECONDS) IN THE TABLE CORRESPOND TO THE MINIMUM CROSS CORRELATION BETWEEN BREATHING SIGNALS OBTAINED BY THE PROPOSED METHOD AND THE REFERENCE METHOD

Subject	Protocol 1	Protocol 3	Protocol 4
1	-0.48	1.56	1.12
2	18.2	3.4	3.16
3	0.36	2.00	1.04
4	-0.36	-0.48	-2.44
5	0.12	2.04	2.24
6	0.40	0.84	0.92

### C. Evaluation of Breathing Cycles and Breathing Rate

To further evaluate the performance of the proposed method, the numbers of breathing cycles as well as the breathing rate were obtained from the estimated and corresponding reference breathing signals of protocols 1 and 3, and they were compared to each other. Detected peaks of estimated and corresponding reference breathing signals are shown in Fig. 12. Linear regression analysis and a Bland–Altman plot were used for the validation of data. The linear regression and Bland–Altman plot analyses of breathing rate are illustrated in Fig. 13. The coefficient of determination of the regression line is 0.34. The mean of differences in the Bland–Altman plot is  $-1.75$  bpm, and 95% limits of the agreement are 6.2716 bpm (upper boundary) and  $-9.7716$  bpm (lower boundary). The scatter plot and regression line of estimated and reference breathing cycles are shown in Fig. 14(a). The coefficient of determination of the regression line is 0.59. An optimal regression line (red dashed line) is also illustrated in Fig. 14(a).

The Bland–Altman plot with respect to the proposed method and reference method is illustrated in Fig. 14. The mean of differences is  $-1.3333$  number of breathing cycles, and 95% limits of agreement are 4.7126 number of breathing cycles (upper boundary) and  $-7.3793$  number of breathing cycles (lower boundary). Only one point is outside the interval.

### D. Evaluation of Breathing Patterns

Table III presents the mean absolute errors (MAEs) and root mean square errors (RMSEs) between the instantaneous frequencies of the estimated breathing pattern sequence and instantaneous frequencies of the corresponding reference breathing patterns. The mean and standard deviation of MAEs are 0.1235 and 0.0201 Hz, respectively, while the mean and standard deviation of RMSEs are 0.1684 and 0.0349 Hz, respectively. Instantaneous frequencies of the estimated and reference breathing signals are visually illustrated in Fig. 15. The estimated and reference breathing patterns of two participants are illustrated in Fig. 16.

## V. DISCUSSION

This study demonstrates that respiration measurement can be done by the joint use of RGB camera and thermal camera, which has a low thermal resolution, with a decent accuracy. The cross-spectrum mapping method extracted well the breathing signal without the need of image registration for simple scenarios but the performance became poor when participants

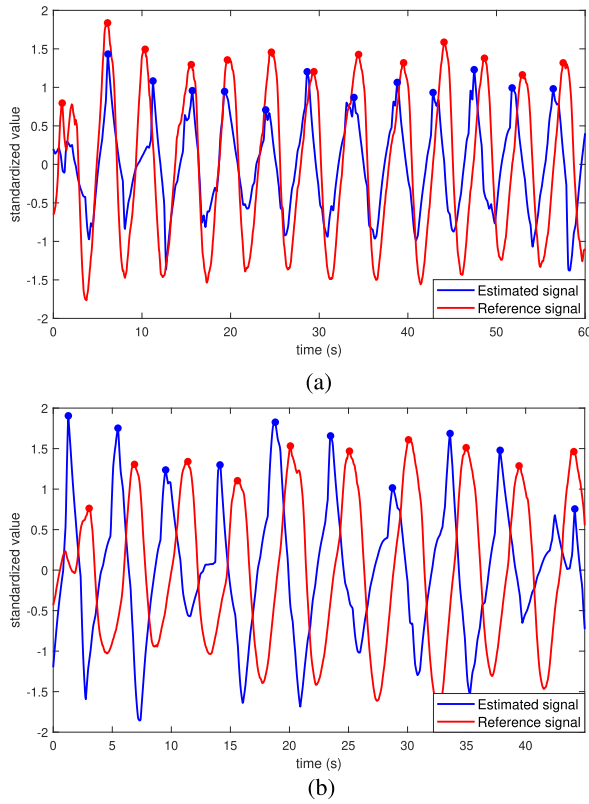


Fig. 12. Number of breathing cycles from breathing signals was computed by detecting peaks of the signals. (a) Protocol 1. (b) Protocol 2.

TABLE III

MAES AND RMSES OF INSTANTANEOUS FREQUENCIES OF PROTOCOL 4 AS WELL AS THEIR MEANS AND STANDARD DEVIATIONS ARE PRESENTED IN THIS TABLE

	MAE	RMSE
1	0.1027	0.1336
2	0.1422	0.1883
3	0.1397	0.1862
4	0.0970	0.1210
5	0.1397	0.2122
6	0.1199	0.1693
<b>mean</b>	<b>0.1235</b>	<b>0.1684</b>
<b>std</b>	<b>0.0201</b>	<b>0.0349</b>

turned their heads around or moved their hands toward the faces. The alignment of the signals was analyzed using linear regression and cross correlation which indicate that there is no strong linear correlation between the estimated and reference breathing signal samples and which also can indicate that the signals are not properly aligned. Breathing rates and the number of breathing cycles were obtained from the estimated and the corresponding reference breathing signals, and they were used for evaluating the proposed method using the linear regression analysis and the Bland–Altman plot. Based on the analyses, the proposed method performed better in computing the number of breathing cycles than in the breathing rate measurement. Based on the MAEs and RMSEs in Table III, our proposed method performed the breathing pattern measurement relatively better for some participants compared to the other participants.

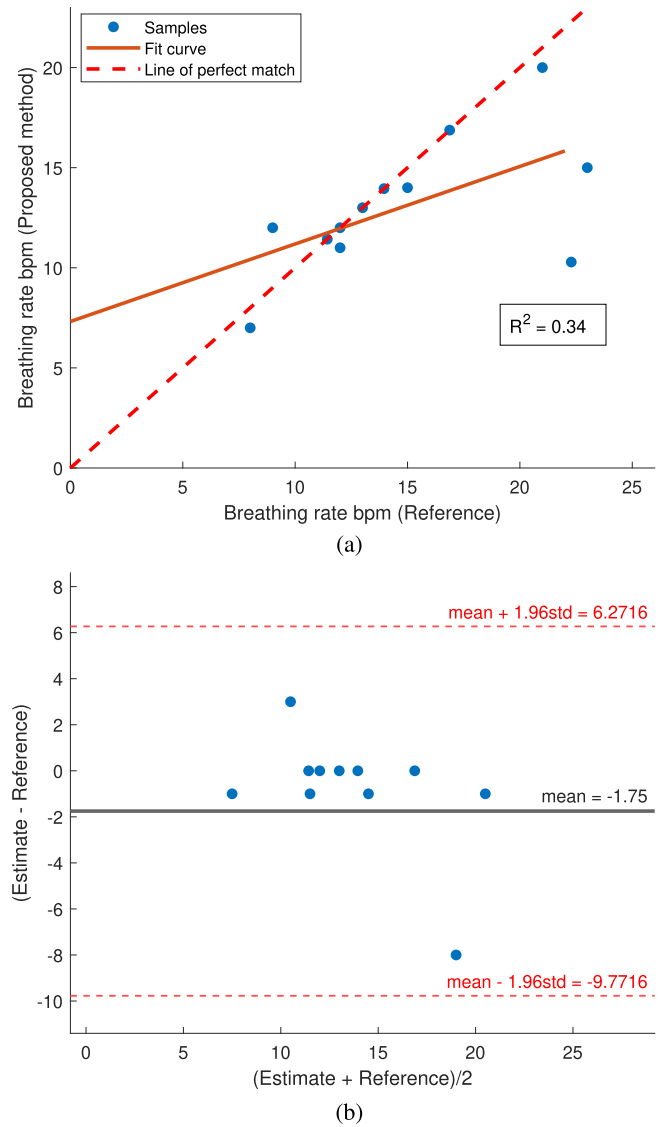


Fig. 13. (a) Correlation between the breathing rate computed from the reference signal and the estimated signal is presented by the fit regression line. The dashed line represents the ideal regression line. (b) Bland–Altman plot compares the estimated and the reference breathing rates. No point is outside the interval.

The cross-spectrum mapping used in the proposed method performed well when there was little to no head movement. Therefore, the performance of the cross-spectrum mapping in protocols 1, 3, and 4 was satisfactory, whereas the performance deteriorated in protocol 2. The cross-spectrum mapping methods in [2], [7], [9], [10], [12], and [13] are relatively more robust to the head and body movements. However, the deterioration of the performance in this work could be due to manual synchronization of videos. Moreover, the frame rate of Android camera was unstable [23], and therefore, the RGB and thermal videos did not possibly maintain synchronization throughout which deteriorated the performance of the cross-spectrum mapping in scenarios corresponding to protocol 2. Furthermore, the tracker used in RGB videos lost the ROI when a participant turned the head to the left or right side which also explains the poor performance of the cross-spectrum mapping in protocol 2.

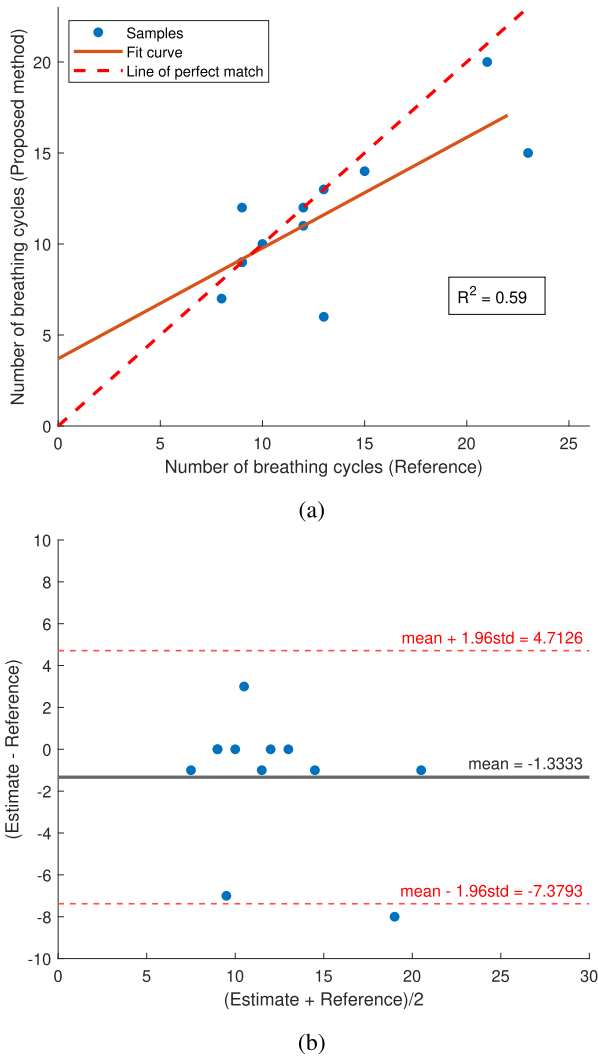


Fig. 14. (a) Correlation between the number of breathing cycles detected from the reference method and the proposed method is presented by the fit regression line. The dashed line represents the ideal regression line. It seems that there are only 11 data points because two points overlap each other. (b) Bland–Altman plot compares the estimated and the reference number of detected breathing cycles. Only one point is outside the interval.

The low  $R^2$  values in Table I indicate that the linear correlation between the estimated and reference breathing signal samples is weak, and therefore, the estimated and reference signals were not properly aligned. However, based on the mere coefficient of determination, it is not necessarily possible to determine goodness of fit because variation in the independent variable affects largely on the coefficient of determination [24]. Another problem was that some of the slopes of the regression lines were positives [see Fig. 11(c)] which is inconsistent with the nostril temperature being inversely proportional to the inductance of the RIP belt. Since the proposed and the reference breathing measurement methods measure two different physical properties that do not necessarily behave in the similar way during the inhalation and exhalation, it is possible that in reality, there is simply no linear correlation between the proposed and the reference methods. Time lags presented in Table II indicate that the signals are not aligned properly. Time lags of subject 2 are relatively large compared

to other subjects. As mentioned before, Android cameras have an unstable frame rate, and therefore, estimated timestamps of video frames deviate from the actual timestamps. One of the reasons for subject 2 having large time lags could be relatively large deviations between the estimated and actual timestamps in comparison to other subjects. One of the reasons for the weak correlation between the estimated signal and the reference signal is the visual estimation of the calibration timestamps in camera-belt calibration, which was used for the signal alignment. Since it is difficult to determine the exact timestamps visually, signals are going to be more or less misaligned. Besides the visual estimation of the calibration timestamps, the unstable frame rates of the cameras are also one of the factors for the weak correlation, because in this work, we assumed the cameras recording at a constant frame rate. The alignment accuracy between the estimated and reference breathing signals has been studied very little in previous related works even though it is important for obtaining reliable results. In [25], the alignment agreement between the signals was evaluated visually.

Fig. 13(a) indicates that the coefficient of determination for breathing rates ( $R^2 = 0.34$ ) is lower than previously obtained coefficients of determination, such as  $R^2 = 0.971$  [7] and  $R^2 = 0.8137$  [9]. Ideally, the coefficient of determination would be  $R^2 = 1$  in our case. Moreover, our mean of differences ( $-1.75$ ) in the Bland–Altman plot in Fig. 13(b) is not as close to the ideal mean of differences (0) as the ones obtained in the previous studies, such as 0.1527 [2] and  $-0.304$  [7]. Based on the linear regression [see Fig. 14(a)] and Bland–Altman plot [see Fig. 14(b)], the proposed method performs better in computing the numbers of breathing cycles ( $R^2 = 0.59$ , mean =  $-1.3333$ ) than computing the breathing rates ( $R^2 = 0.34$ , mean =  $-1.75$ ). However, computing only the number of breathing cycles is not clinically meaningful. There are many reasons for the relatively low  $R^2$  value for the breathing rate measurement. First of all, number of participants in our experiment is small for generalizing the performance of our proposed method. Second, as mentioned before, variation in independent variable affects largely on the coefficient of determination. Therefore, the  $R$  value is not the best way to evaluate the performance of the proposed method, especially when the sample size is small. Also, there were a few cases where some of the breathing peaks were not detected, because the signal processing might suppress those peaks below the threshold level. Regardless, the low thermal resolution thermal camera has the potential for measuring the breathing rate.

The deviation metrics (MAE and RMSE) between the estimated and reference breathing pattern sequence is comparable to the previous study. Ideally, MAE and RMSE values presented in Table III would be zeros which would indicate that there is no deviation at all between the estimated and the reference breathing pattern sequences. Maurya et al. [12] evaluated their proposed RGB-thermal image registration-based breathing monitoring method by computing the MAE and RMSE between the instantaneous respiration rates obtained by their proposed and reference methods. According to Maurya et al., the mean MAE and mean RMSE for the study, where participants breathe in various patterns, are 2.0786 and 3.4172,

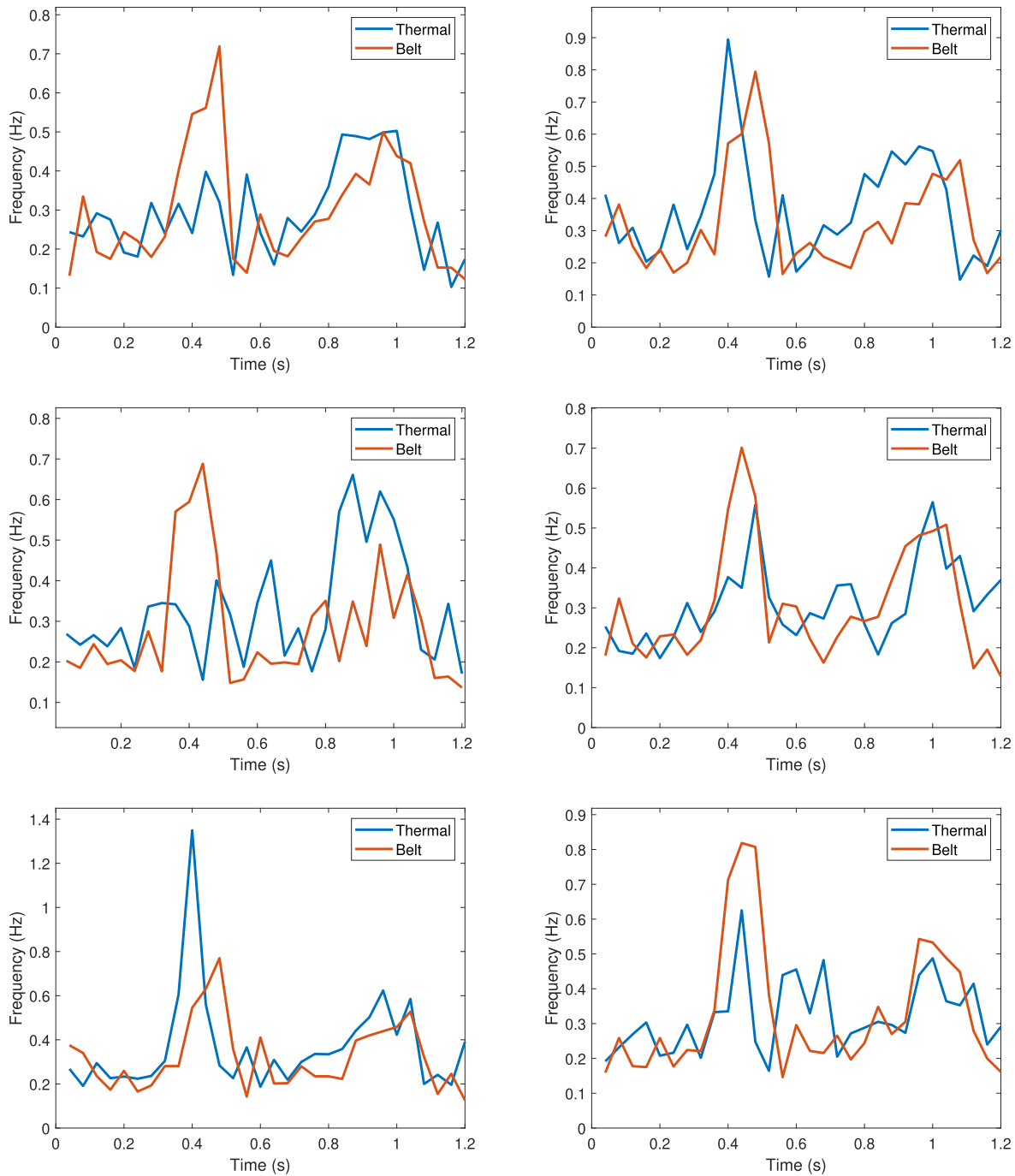


Fig. 15. Comparisons of instantaneous frequencies of estimated breathing signals (blue) and reference breathing signals (orange) that were measured during protocol 4 are visually illustrated. Each chart represents an individual subject.

respectively, whereas the mean MAE and mean RMSE in this work are lower (0.1235 and 0.1684, respectively). This suggests that in our work, the estimated breathing pattern sequence deviates less from the reference breathing pattern than in [12]. Therefore, it can be inferred that our proposed method improves the breathing pattern sequence measurement. Although good MAE and RMSE values were obtained, it is not feasible to use visual illustrations of breathing patterns (see Figs. 15 and 16) to demonstrate that our method can correctly determine the breathing patterns. For example, the apnea phases are unnoticeable in Fig. 16 which could be a

result of the signal processing. Therefore, quantitative analysis is important in this case. Regardless, there is a potential for a thermal camera with a low thermal resolution to measure different breathing patterns with a decent level of accuracy.

The calibration process and time synchronization have some improvement possibilities. Speech-based time synchronization may not be an accurate method because each subject spoke with a different sound level in this work. Because of this, a silence threshold, which is an upper bound for how quiet is silent, was adjusted for each subject. In future works, sound-based time synchronization between two videos can be



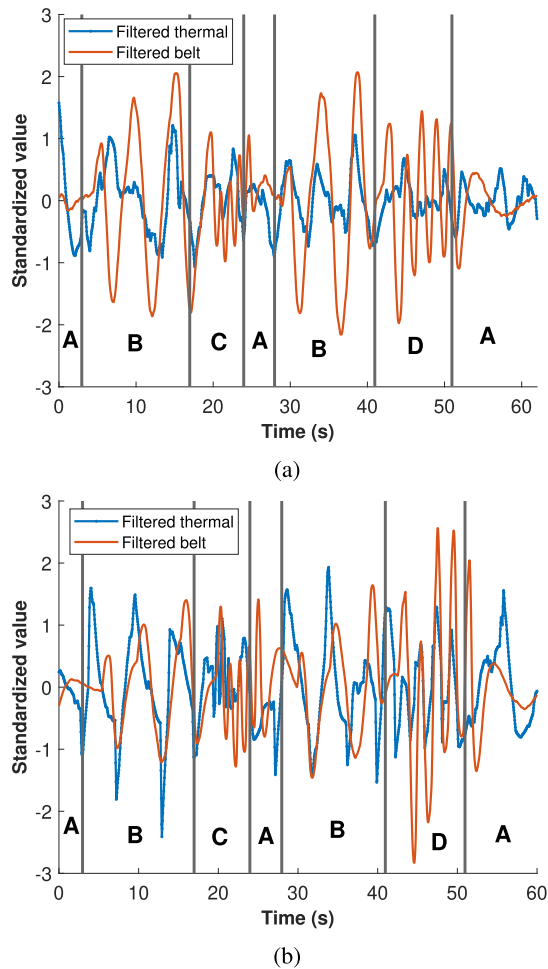


Fig. 16. Estimated signal (blue) and corresponding reference signal contain the following breathing patterns: apnea (A), eupnea (B), tachypnea (C), and Kussmaul (D). Each segment (denoted by a letter) corresponds to a breathing pattern. Apnea is not clearly visible due to noise in signal. (a) Subject 1. (b) Subject 2.

implemented, for example by using the same sound source that produces the same sound pattern for each subject so that the sound pattern can be extracted from both RGB video and thermal video. Thus, the time synchronization would become more standardized and accurate.

The calibration process was performed manually (i.e., the calibration points were selected manually) in this work, which is not an accurate method. Human error in selecting the calibration points is one of the sources of the calibration inaccuracy. Another error source is the used material of the calibration object. For the RGB-FLIR RGB camera calibration, paper with grids marked with a marker was used, whereas for the FLIR RGB-thermal camera calibration, paper with square holes. In future work, the selection of calibration points should be done by an automatic algorithm.

## VI. CONCLUSION

In this article, we have presented a method for respiration measurement using optical (RGB) and low thermal resolution thermal cameras. Previous works employed high thermal resolution thermal cameras, which are expensive as well as large and bulky making them less portable. Conversely, the

low thermal resolution thermal camera used in this work was small in size and easily portable. In this method, the cameras are calibrated and synchronized using calibration objects and sound analysis tools, respectively, in order to perform cross-spectrum mapping for tracking ROI in thermal videos. The study shows that cross-spectrum mapping performs well when there is little to no head movement. The respiration measurement is done by measuring the temperature changes inside of the nostrils using the thermal camera. The linear regression analysis and cross correlation show that there is no strong linear correlation between the estimated and reference signals, and the signals were not perfectly aligned. This is most probably due to not use of exact timestamps because they were inaccessible. Although the proposed method did not measure the breathing rate as accurately as in the previous studies, where thermal cameras with higher thermal resolutions were used, the proposed method is a promising breathing measurement method. The breathing pattern measurement results show that the proposed method has the potential to capture different breathing patterns.

## REFERENCES

- [1] O. Steichen, G. Grateau, and E. Bouvard, "Respiratory rate: The neglected vital sign," *Med. J. Aust.*, vol. 189, no. 9, pp. 531–532, Nov. 2008.
- [2] L. Chen, M. Hu, N. Liu, G. Zhai, and S. X. Yang, "Collaborative use of RGB and thermal imaging for remote breathing rate measurement under realistic conditions," *Infr. Phys. Technol.*, vol. 111, Dec. 2020, Art. no. 103504.
- [3] J. Amit Benjamin and K. Lewis, "Sleep-disordered breathing and cardiovascular disease," *Postgraduate Med. J.*, vol. 84, no. 987, pp. 15–22, 2008.
- [4] R. Heinzer et al., "Prevalence of sleep-disordered breathing in the general population: The HypnoLaus study," *Lancet Respiratory Med.*, vol. 3, no. 4, pp. 310–318, Apr. 2015.
- [5] C. Massaroni, A. Nicolo, D. Lo Presti, M. Sacchetti, S. Silvestri, and E. Schena, "Contact-based methods for measuring respiratory rate," *Sensors*, vol. 19, no. 4, p. 908, Feb. 2019.
- [6] S. Aario, A. Gorad, M. Arvonen, and S. Särkkä, "Respiratory pattern recognition from low-resolution thermal imaging," in *Proc. ESANN*, Belgium, Europe, 2020, pp. 469–474.
- [7] M.-H. Hu, G.-T. Zhai, D. Li, Y.-Z. Fan, X.-H. Chen, and X.-K. Yang, "Synergetic use of thermal and visible imaging techniques for contactless and unobtrusive breathing measurement," *J. Biomed. Opt.*, vol. 22, no. 3, p. 1, Mar. 2017, Art. no. 036006.
- [8] C. Massaroni, A. Nicolo, M. Sacchetti, and E. Schena, "Contactless methods for measuring respiratory rate: A review," *IEEE Sensors J.*, vol. 21, no. 11, pp. 12821–12839, Jun. 2021.
- [9] L. Chen, N. Liu, M. Hu, and G. Zhai, "RGB-thermal imaging system collaborated with marker tracking for remote breathing rate measurement," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2019, pp. 1–4.
- [10] L. Maurya, P. Mahapatra, and D. Chawla, "Simultaneous breathing monitoring of multiple persons using thermal and visible imaging," *IEEE Sensors J.*, vol. 21, no. 24, pp. 28057–28065, Dec. 2021.
- [11] T. Negishi et al., "Contactless vital signs measurement system using RGB-thermal image sensors and its clinical screening test on patients with seasonal influenza," *Sensors*, vol. 20, no. 8, p. 2171, Apr. 2020.
- [12] L. Maurya, P. Mahapatra, and D. Chawla, "Non-contact breathing monitoring by integrating RGB and thermal imaging via RGB-thermal image registration," *Biocybernetics Biomed. Eng.*, vol. 41, no. 3, pp. 1107–1122, Jul. 2021.
- [13] T. Negishi, G. Sun, H. Liu, S. Sato, T. Matsui, and T. Kirimoto, "Stable contactless sensing of vital signs using RGB-thermal image fusion system with facial tracking for infection screening," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 4371–4374.
- [14] F. Deng et al., "Design and implementation of a noncontact sleep monitoring system using infrared cameras and motion sensor," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 7, pp. 1555–1563, Jul. 2018.

- [15] I. Lorato et al., "Multi-camera infrared thermography for infant respiration monitoring," *Biomed. Opt. Exp.*, vol. 11, no. 9, p. 4848, 2020.
- [16] FLIR Systems. (2017). *FLIR One Pro 3rd Generation User's Guide*. Accessed: Jun. 2, 2023. [Online]. Available: <https://www.flir.com/globalassets/imported-assets/document/flir-one-pro-user-guide-3rd-gen-en.pdf>
- [17] J. Robert and M. Webbie. (2018). *Pydub*. Accessed: Jun. 2, 2023. [Online]. Available: <http://pydub.com/>
- [18] G. Bradski, "The OpenCV library," *Dr. Dobb's J. Softw. Tools*, vol. 120, no. 11, pp. 122–125, 2000.
- [19] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1867–1874.
- [20] A. Lukežić, T. Vojt, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4847–4856.
- [21] F. Q. Al-Khalidi, R. Saatchi, D. Burke, and H. Elphick, "Tracking human face features in thermal images for respiration monitoring," in *Proc. ACS/IEEE Int. Conf. Comput. Syst. Appl.*, May 2010, pp. 1–6.
- [22] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2008.
- [23] F. Guede-Fernández, V. Ferrer-Mileo, J. Ramos-Castro, M. Fernández-Chimeno, and M. A. García-González, "Real time heart rate variability assessment from Android smartphone camera photoplethysmography: Postural and device influences," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 7332–7335.
- [24] D. F. Hamilton, M. Ghert, and A. H. R. W. Simpson, "Interpreting regression models in clinical outcome studies," *Bone Joint Res.*, vol. 4, no. 9, pp. 152–153, Sep. 2015.
- [25] C. Barbosa Pereira, X. Yu, M. Czaplik, V. Blazek, B. Venema, and S. Leonhardt, "Estimation of breathing rate in thermal imaging videos: A pilot study on healthy human subjects," *J. Clin. Monitor. Comput.*, vol. 31, no. 6, pp. 1241–1254, Dec. 2017.

**Zaeed Khan** received the B.Sc. (Tech.) degree in bio-information technology and the M.Sc. (Tech.) degree in biosensing and bioelectronics from Aalto University, Espoo, Finland, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Automation.

He joined the Sensor Informatics and Medical Technology Research Group, Aalto University, in 2021. His research interests include vital sign monitoring, computer vision, and machine learning for classification and prediction in the domain of health care.



**Matias Rusanen** received the B.Sc. and M.Sc. degrees in medical physics from the University of Eastern Finland, Kuopio, Finland, in 2019 and 2021, respectively. He is currently pursuing the Ph.D. degree with the Department of Technical Physics, University of Eastern Finland, and the Diagnostic Imaging Center, Kuopio University Hospital, Kuopio.

He joined the Sleep Technology and Analytics Research Group, University of Eastern Finland, in 2019. His research interests include the development and validation of wearable sensor technology for diagnostics of sleep disorders and the integration of deep learning methods with these systems.

**Miika Arvonen** currently serves as a Consultant for pediatric infectious diseases at Kuopio University Hospital, Kuopio, Finland. His research is dedicated to exploring infectious diseases, immunology, and respiratory measurement.



**Timo Leppänen** received the B.Sc., M.Sc., and Ph.D. degrees in medical physics from the University of Eastern Finland, Kuopio, Finland, in 2014, 2015, and 2016, respectively.

He started his studies with the University of Eastern Finland, in 2011, after two years of service with the Finnish Defense Forces. Since 2017, he has been a Principal Investigator and the Co-Head of the Sleep Technology and Analytics Research Group, University of Eastern Finland, where he was appointed as an Associate Professor in 2022. His research

interests include the development of artificial intelligence solutions for diagnostics and severity estimation of sleep apnea, phenotyping and progression of obstructive sleep apnea, sleep apnea-related psychomotor vigilance and daytime sleepiness, and the development of wearable sensors for home sleep apnea testing.

Dr. Leppänen is a Board Member of the European Sleep Research Society's Scientific Committee and the Finnish Sleep Research Society. In addition, he is a member of the Finnish Society for Medical Physics and Medical Engineering and the Finnish Society of Clinical Neurophysiology.



**Simo Särkkä** (Senior Member, IEEE) received the M.S. (Tech.) degree (Hons.) in engineering physics and mathematics and the D.Sc. (Tech.) degree (Hons.) in electrical and communications engineering from the Helsinki University of Technology, Espoo, Finland, in 2000 and 2006, respectively.

He is currently an Associate Professor with Aalto University, Espoo, and an Adjunct Professor with the Tampere University of Technology, Tampere, Finland, and the Lappeenranta University of Technology, Lappeenranta, Finland. He has authored or

coauthored over 150 peer-reviewed scientific articles and the books *Bayesian Filtering and Smoothing* (Cambridge University Press) and *Applied Stochastic Differential Equations* (Cambridge University Press) along with the Chinese translation of the former. His research interests include multisensor data processing systems with applications in location sensing, health and medical technology, machine learning, inverse problems, and brain imaging.

Dr. Särkkä serves as a Senior Area Editor for IEEE SIGNAL PROCESSING LETTERS.