

# Privacy-Preserved Contactless Sleep Parameters Measurement Using a Defocused Camera

Yingen Zhu<sup>1</sup>, Hong Hong<sup>1</sup>, Senior Member, IEEE, and Wenjin Wang<sup>1</sup>

**Abstract**—Sleep monitoring plays a vital role in various scenarios such as hospitals and living-assisted homes, contributing to the prevention of sleep accidents as well as the assessment of sleep health. Contactless camera-based sleep monitoring is promising due to its user-friendly nature and rich visual semantics. However, the privacy concern of video cameras limits their applications in sleep monitoring. In this paper, we explored the opportunity of using a defocused camera that does not allow identification of the monitored subject when measuring sleep-related parameters, as face detection and recognition are impossible on optically blurred images. We proposed a novel privacy-protected sleep parameters measurement framework, including a physiological measurement branch and a semantic analysis branch based on ResNet-18. Four important sleep parameters are measured: heart rate (HR), respiration rate (RR), sleep posture, and movement. The results of HR, RR, and movement have strong correlations with the reference (HR:  $R = 0.9076$ ; RR:  $R = 0.9734$ ; Movement:  $R = 0.9946$ ). The overall mean absolute errors (MAE) for HR and RR are 5.2 bpm and 1.5 bpm respectively. The measurement of HR and RR achieve reliable estimation coverage of 72.1% and 93.6%, respectively. The sleep posture detection achieves an overall accuracy of 94.5%. Experimental results show that the defocused camera is promising for sleep monitoring as it fundamentally eliminates the privacy issue while still allowing the measurement of multiple parameters that are essential for sleep health informatics.

Manuscript received 20 December 2023; revised 11 March 2024; accepted 27 April 2024. Date of publication 2 May 2024; date of current version 7 August 2024. This work was supported in part by the National Key R&D Program of China under Grant 2022YFC2407800, in part by the National Natural Science Foundation of China under Grant 62271241 and Grant 62350068, in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515012983, in part by Shenzhen Science and Technology Program under Grant JSGGKQTD20221103174704003, and in part by Shenzhen Fundamental Research Program under Grant JCYJ20220530112601003. (*Corresponding authors:* Hong Hong; Wenjin Wang.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Institutional Review Board of Southern University of Science and Technology under Application No. 2023131, and performed in line with the GCP standards.

Yingen Zhu and Wenjin Wang are with the Department of Biomedical Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: wangwj3@sustech.edu.cn).

Hong Hong is with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: hongnu@njust.edu.cn).

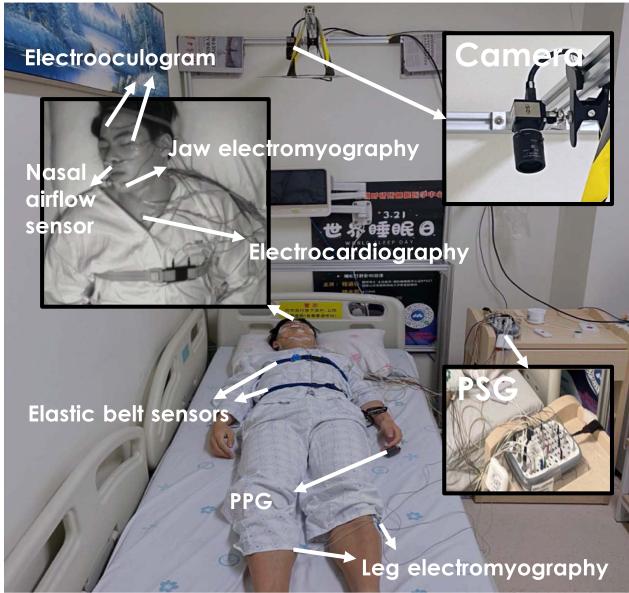
Digital Object Identifier 10.1109/JBHI.2024.3396397

**Index Terms**—Contactless sensing, defocused camera, privacy protection, sleep monitoring.

## I. INTRODUCTION

SLEEP monitoring is vital for understanding sleep quality and personal health, playing a crucial role in multiple settings (e.g., NICU: Neonatal Intensive Care Unit, ICU: Intensive Care Unit, sleep centers, postpartum centers, elderly care, baby care, home monitoring). In NICU and ICU settings, continuous monitoring of vital signs (e.g., RR: respiratory rate, HR: heart rate, SpO<sub>2</sub>: oxygen saturation, BP: blood pressure), especially during sleep, has been commonly used for detecting abnormalities and alerting clinicians to reduce the incidence of adverse events in premature infants and high-risk postoperative patients [1]. In other settings, sleep monitoring extends beyond accident prevention to assess a person's health informatics [2]. The continuous monitoring the aforementioned vital signs, as well as sleep posture, wake-ups, and sleep motility allows a comprehensive evaluation of sleep quality and it can determine if a person has sufficient and restorative sleep to minimize the risk of postpartum depression [3]. It can also identify sleep disorders (e.g., sleep apnea, insomnia) and provide early warnings for potential health issues (e.g., obesity, arrhythmia, hypertension) for better medication or healthcare advice [4]. The physiological data of sleep from these settings can support the treatment of sleep disorders, leading to improved therapies [5]. Hence, fully automatic continuous sleep monitoring holds promising application values in various use cases.

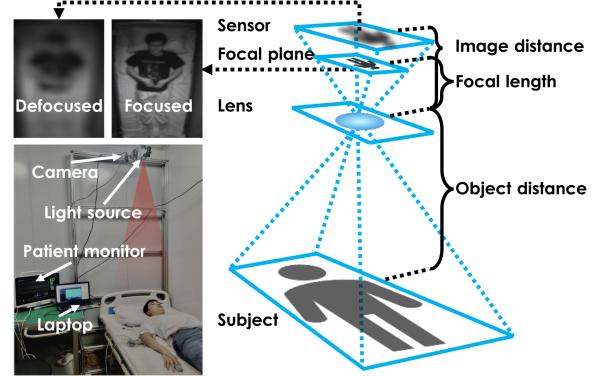
Currently, the gold standard for sleep monitoring in a clinical environment is Polysomnography (PSG, see Fig. 1) under the supervision of professional clinicians (sleep experts and technicians). The application of PSG requires biomedical sensors attached to the patient body, which limits the mobility of the patient and poses significant risks of contagious infections between patients and clinicians [6]. In the sleep center, the commonly used PSG system (Philips Respironics ALICE 6 LDxN, Version 3.9.5) has a total of 54 cables that need to be connected to the patient body, which is expensive, patient-unfriendly, and complex for long-term use, triggering the exploration of more comfortable and convenient alternatives. The most promising option is using less intrusive options instead of conventional contact-based sensors, such as Radio Frequency (RF) equipment, depth sensors, thermal cameras, and optical cameras.



**Fig. 1.** Comparison of PSG and contactless camera setups for sleep monitoring in a hospital sleep center.

Among them, optical camera-based sleep parameters measurement is more favored due to its properties of low-cost and easy of access, and rich visual semantics that are useful for measuring various sleep parameters (e.g., sleep posture, body actigraphy, and bed-exiting/falling) [7].

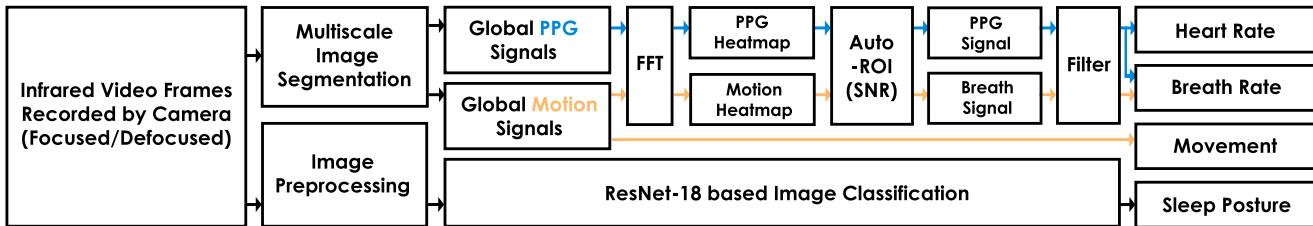
For camera-based sleep monitoring, it enables the measurement of two main types of signals: physiological signals and contextual signals. The abundant information of a video is exploited for measuring contexts such as sleep postures [8] and emotions [9], where remote photoplethysmography (rPPG) has been used for measuring vital signs (HR and RR) during sleep. More specifically, face is an important Region of Interest (RoI) for video-based measurement during sleep, which has been used for pulse extraction and posture estimation [10]. However, it poses significant privacy concerns as it exposes sensitive information like identity, particularly in privacy-sensitive settings such as home-based sleep monitoring. The introduction of regulations such as the EU's General Data Protection Regulation (GDPR) [11] and US' California Consumer Privacy Act [12], emphasizes that personal data including images cannot be collected without users' consent in any situation, which inevitably constrains the commercialization of camera-based sleep monitoring. Although GDPR allows personal data to be processed for medical purposes, it still depends on the agreement between patients and clinical/business units, especially in privacy-sensitive medical applications. Besides, consumers are hesitant to embrace such systems in private settings due to the fears of leakage and misuse of personal data [13]. Although privacy may be less of a concern in hospitals or sleep centers as compared with homes, the inpatients sensitive to privacy can opt to have their identity information protected. Therefore, for camera-based sleep monitoring, addressing privacy concerns is crucial for achieving ubiquitous applications and general popularity.



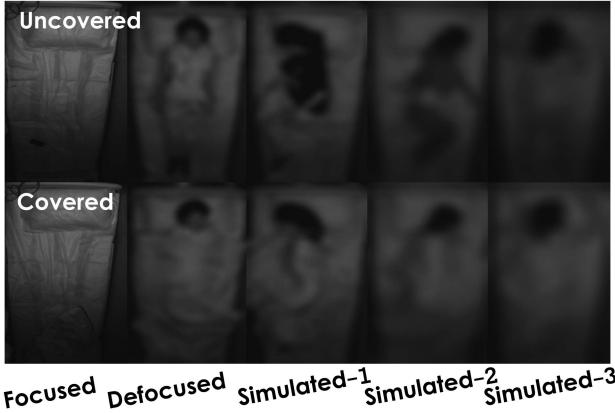
**Fig. 2.** A sleep monitoring setting simulated in the dark chamber and the principle of how the camera is defocused to obtain blurry images.

To address this issue, an earlier study proposed to eliminate the privacy issue using a single-element camera [14], but it sacrifices important spatial context information required for semantic measurement, i.e., the single-element approach cannot measure motion signals (e.g., respiratory signal) nor sleep postures. This paper attempts to explore a compromised option between the single-element sensor (e.g., photodiode) and high-resolution camera, i.e. the defocused camera that eliminates the primary concern of privacy protection, which is the de-identification in sleep videos. This can be referred to Figs. 2 and 5, where face detection and recognition is impossible for both human and machine but still allows the processing of certain spatial contexts, e.g., motion analysis, posture estimation, and in-bed detection. For a comprehensive sleep parameters measurement, this study aims to detect four sleep parameters, including two critical vital signs (HR and RR), sleep posture, and movement. For HR and RR extraction, we extract pulse and respiratory signals from the image sequence using a spatial-redundant framework [15] that integrates a fully-automatic living-skin detection approach for RoI localization [16]. Note that the respiration can be extracted from pixel movement and low-frequency modulation in the PPG signal. Hence, we denote them by motion-based RR and PPG-based RR, respectively. When a camera is in the defocused condition, the benchmark shows that both the HR measurement and PPG-based RR measurement have high correlations with the reference ( $HR:R = 0.9076$ ;  $RR:R = 0.9734$ ), with an overall mean absolute error (MAE) of 5.2 bpm (beats per minute) for HR and 1.5 bpm for RR over 19 subjects, respectively. The measurement of HR and RR achieve reliable estimation coverage of 72.1% and 93.6%, respectively. Movement measurement in blurry images also has a high agreement with the reference ( $R = 0.9946$ ). For sleep posture recognition, we retrain the ResNet-18 network on the augmented dataset based on our sleep recordings. It achieves effective recognition of five states (off-bed/empty, supine, left-side, right-side, prone), with an overall accuracy of 94.5% in our benchmark.

This paper mainly contributes to the following aspects: it (i) proves that the defocused camera is a viable solution for protecting privacy in sleep monitoring; (ii) shows the feasibility of measuring HR, RR, motion, and sleep postures from videos



**Fig. 3.** The video processing framework for extracting sleep parameters from focused/defocused videos, with the blue line representing the processing of PPG signals and the yellow line representing the processing of motion signals.



**Fig. 4.** From left to right, snapshots of five degrees of defocus are exemplified. Each video condition includes five different sleep postures (off-bed/empty, supine, left-side, right-side, prone), and the figure only shows one sleep posture for each condition. The first row shows postures without bedsheets occlusion, while the second row shows postures with bedsheets occlusion.

recorded by a defocused camera; (iii) carries out a thorough investigation on the impact of bedsheets occlusions, sleep postures, and degrees of image blur, which increases the understanding of the value of a defocused camera for sleep assessment.

The rest of this paper is organized as follows. Section II provides an overview of the related work about contactless physiological measurement, especially the camera-based monitoring and its privacy concerns. Section III details the proposed sleep monitoring framework with a defocused camera. The experimental setup is discussed in Section IV while the results are discussed in Section V. Finally, the conclusions are drawn in Section VI.

## II. RELATED WORK

This section comprises two parts. The first part discusses the major techniques used for contactless physiological measurement, and the second part focuses on privacy protection for sleep monitoring. Table I shows the feasibility of measuring six physiological parameters for five different devices.

### A. Contactless Physiological Measurement

Various contactless methods for physiological measurement have been proposed based on different hardware configurations, including RF equipment (e.g., WiFi, ultra-wideband radar, FMCW radar), depth sensors, microphones, thermal and optical

**TABLE I**  
THE FEASIBILITY OF MEASURING PHYSIOLOGICAL PARAMETERS USING DIFFERENT DEVICES

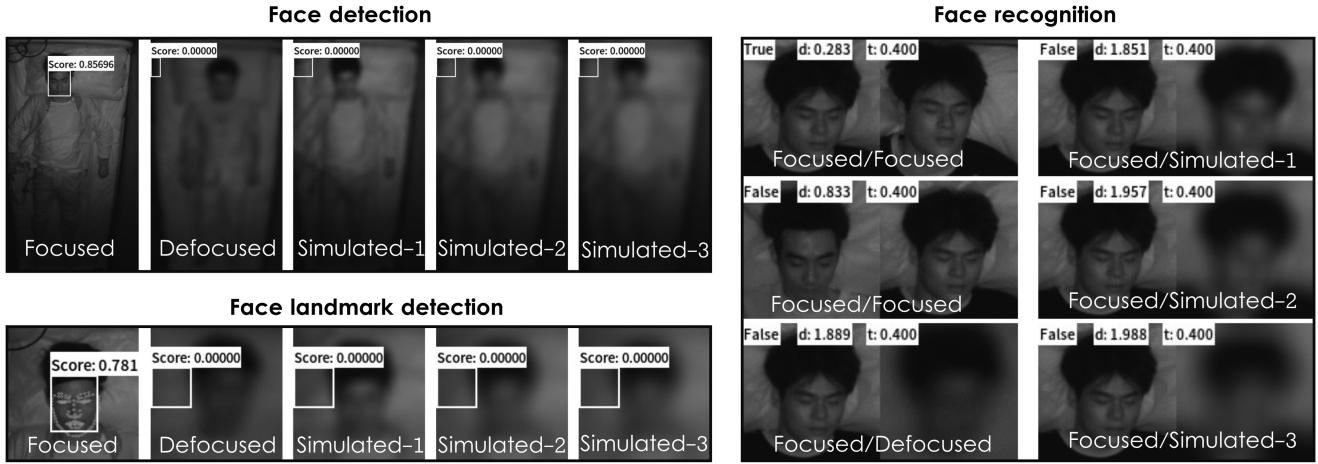
Devices	Physiological parameters					
	RR	HR	SP	M	SpO <sub>2</sub>	BP
RF equipment	✓	✓	✓	✓	✗	✗
Depth sensor	✓	✗	✓	✓	✗	✗
Microphone	✓	✗	✗	✗	✗	✗
Thermal camera	✓	✗	✓	✓	✗	✗
Optical camera	✓	✓	✓	✓	✓	✓

\*“RR”, “HR”, “SP”, “M”, “SpO<sub>2</sub>”, and “BP” refer to six sleep parameters , namely respiration rate, heart rate, sleep posture, movement,oxygen saturation, and blood pressure.

cameras (e.g., RGB, infrared). The following discusses different devices used for physiological measurement during sleep.

RF equipment emits electromagnetic waves and senses the environment using the Doppler effect, which refers to the frequency alteration of a signal caused by relative movement between transmitter and receiver [17]. The non-invasive and penetrative nature of electromagnetic waves gathers valuable health-related information while protecting privacy. This empowers RF devices to conduct non-line-of-sight monitoring and be able to penetrate through smoke and concrete walls [18]. In sleep monitoring, parameters like RR, HR, and sleep posture have been successfully extracted from the echo signals containing information about the subject’s movements (chest and abdominal movements, rollover movement) [19]. Although large motion detection is feasible for RF equipment, micro-motion analysis such as heartbeat extraction is still challenging, and it cannot measure blood absorption information [20]. When radar is used in wild settings, it often requires additional calibration especially when being positioned in a new space with different spatial structures [21].

The depth sensor estimates object distance by emitting signals and measuring their round-trip traveling time, which allows the measurement of features with clear spatial position changes (RR and sleep posture) [22]. The commercial depth sensors can achieve the best accuracy of  $\pm 1$  cm at the distance of 2 m [23]. This single coarse-grained modality makes it difficult to accurately measure other sleep parameters such as HR [24]. The microphone can only detect acoustic waves of breathing caused by the air pressure changes during sleep [25], and its limited frequency response range cannot capture weaker signals like heartbeats or muscle movements, nor blood perfusion indices such as BP and SpO<sub>2</sub>. For the thermal camera, the



**Fig. 5.** Assess privacy protection effectiveness using three common face analysis tasks under five defocusing conditions. (1) Face detection: White box marks detected face positions with confidence numbers; If no face is detected, it is placed at the top-left corner with confidence set to **0**. (2) Facial landmark detection: Adding facial landmark detection to the faces marked by white boxes in task 1. (3) Face recognition: Matching two images containing faces. **d** represents cosine similarity, where smaller values indicate more similar faces. **t** represents the threshold for **d**, if it is less than the threshold, they are considered the same face, labeled as **true**; Otherwise, as **false**.

obvious advantages are the measurement of body temperature and its independence of illumination conditions [26]. The RR and sleep posture can be detected by thermal cameras [27], but unfortunately, they cannot measure HR as the wavelength range of thermal cannot measure blood perfusion. The thermal camera has been combined with an optical camera for HR measurement [28]. This not only damages the advantage of thermal cameras in privacy protection but also increases hardware costs.

Based on the optical camera, HR and RR can be measured through the rPPG technique, an optical method that detects subtle color changes of the skin induced by blood pulsation [29]. Besides, motion-based RR, sleep posture, and body movements can also be measured by video processing [30], [31]. Cameras also enable the remote measurement of other physiological parameters (e.g., BP [32], SpO<sub>2</sub> [33]), which is a clear advantage as compared to other reviewed options. By leveraging computer vision techniques, cameras can be used to analyze human behaviors, activities, emotions, etc. [34], [35]. However, cameras invade personal information such as identity, resulting in significant privacy concerns [36] that may deteriorate the application of vision-based sleep monitoring technologies.

For joint sleep monitoring involving multiple contactless devices, Chung et al. measured movement and sound for sleep staging by fusing radar and microphone [37], but blood absorption information like SpO<sub>2</sub> cannot be measured. Other studies fusing optical cameras have measured multiple parameters, including SpO<sub>2</sub>, but the privacy issue was not addressed neither [38]. Comprehensive sleep parameters measurement provides valuable information for the prevention of sleep disorders, follow-up treatment, and overall health assessment. Earlier studies have demonstrated the feasibility of sleep staging through HR, RR, and motion [39]. SpO<sub>2</sub> and sleep posture are important measures for patients with obstructive sleep apnea (OSA) since the proper sleep posture prevents oxygen desaturation [40]. The heart rate variability (HRV) has also been applied to understand autonomic changes during different sleep stages [41]. To this end,

we should monitor sleep-related indices as much as possible. Among various contactless monitoring methods, only RF-based and optical camera-based methods can detect multiple sleep parameters (e.g., HR, RR, sleep posture, and movement) with a single sensory hardware. Apart from privacy concerns, the optical camera-based method has a large potential for pervasive healthcare (i.e., SpO<sub>2</sub> and BP can be optically measured as well) as shown in Table I. For contactless sleep monitoring, especially in private settings, mitigating privacy concerns will make the camera a simplified and elegant solution.

### B. Privacy-Protected Camera-Based Monitoring

The most sensitive data during sleep is the exposed face that has been used for identification/authentication [42]. Additionally, face is the most frequently used ROI for physiological measurement, typically the vital signs related to blood perfusion that need to be measured from living skin-tissues [10].

To protect identity privacy, two categories of methods that eliminate facial features were proposed: synthesis and perturbation. Synthesis methods such as face encryption [43] protect privacy by replacing sensitive facial features in the original image with privacy-preserving content (e.g. phantom or mask) that is predefined or generated based on essential attributes. Perturbation methods such as face morphing [44], incorporate precisely crafted distortion (e.g., noise, nonlinear mapping, and pixel shuffling) into the original image to disrupt sensitive facial features, in order to prevent the leakage of privacy. However, the above methods rely on complex adversarial training on visual data, and such perturbation or modification may disturb the measurement of physiological signals [45]. Besides, they are all digital image processing approaches, which are not physically robust to privacy invasion especially when hackers hack into the camera system and steal the raw data from the sensor before image perturbation or synthesis, i.e., the private information may still be stolen [46]. For instance, although Gupta et al.

reduced identity information of standard-definition raw video by cropping (only partial face is retained), pixel shuffling, and blurring [47], and then extracting the rPPG signals from the processed video stream, it still poses a risk of identity leakage especially when the camera firmware is hacked and the raw video stream is obtained from the ISP processing [48]. In contrast to identity protection, some works focus on protecting physiological signals while exposing the facial features for identification [49], but it cannot address the most sensitive issue of face identification in home-based sleep monitoring. Therefore, it is necessary to explore the methods that can physically/optically eliminate the privacy issue while still allowing the measurement of vital signs to some extent.

More specifically, for sleep monitoring in private settings, we emphasize the importance of privacy protection, considering that privacy protection and measurement of sleep parameters are equally important. A conceptual single-element camera was proposed to eliminate the privacy issue [14], but it sacrifices important spatial information required for semantic measurement, i.e., it can only measure HR, but not body movement nor sleep posture that requires the analysis of spatial context.

From another perspective, there is no risk of identity leakage if input sources have no identifiable information for the monitored subject. Hence, as shown in Fig. 2, we propose to adjust the focal length of the camera to the defocused mode such that the monitored subject is optically blurred in the image. The defocus here is defined as the situation in which image distance is unequal to the focal length [50]. We first verified that the defocused camera is indeed not possible for identification, and then proposed a video processing framework that measures four body signs that are important for sleep analysis. The defocused camera can be regarded as a middle-ground solution between the single-element camera (i.e., photo-diode) and the regular focused cameras, with the advantages of privacy protection and spatial context analysis to some degree. It essentially prioritizes the privacy protection by physically prohibiting the invasion opportunity, thereby fundamentally protecting sensitive information from being recorded or accessed. For our approach, we stress that no other modifications are needed for the hardware apart from adjusting the focal length of the camera.

### III. METHODS

Fig. 2 introduces the principle of how to defocus a camera. Fig. 3 illustrates the video processing pipeline for extracting four sleep-related parameters. Since HR, RR, and movement are all derived from the physiological measurement branch in Fig. 3, we will discuss the extraction of these three parameters together. For the estimation of sleep posture, it is introduced separately since it is derived from the semantic analysis branch.

#### A. Defocused Camera

The defocus procedure is illustrated in Fig. 2, wherein it is important to note that in most real-world situations, the object distance  $d_{object}$  is significantly larger than the image distance  $d_{image}$ . Specifically, in our setup, the  $d_{object}$  is 1.7 meters, and the  $d_{image}$  is 10 millimeters. The relationship between  $d_{object}$

and  $d_{image}$  is defined by the thin lens theory [50]:

$$\frac{1}{f} = \frac{1}{d_{image}} + \frac{1}{d_{object}}, \quad (1)$$

where  $f$  is the focal length;  $d_{image}$  is often considered approximately equal to  $f$ . Keeping other camera parameters constant and adjusting the focal length, we can blur the image when  $f$  is not equal to  $d_{image}$ . A shorter focal length is preferred for a wider field of view. Here we reduced  $f$  from 10 millimeters to 8 millimeters, as a longer focal length may not capture the entire body. With a fixed focus and aperture, we intentionally shorten the focal length to prevent the filming of facial details. Thus, we can quantify the degree of defocusing by calculating the ratio of  $d_{image}$  to  $f$  when  $d_{object}$  is 1.7 meters, which is approximately 1.25. The performance of semantic analysis is largely degraded but still possible when introducing image blur, which will be detailed in the following sections.

#### B. Heart Rate, Respiration Rate and Movement

Machine learning-based face detection techniques failed with the defocused camera. For vital signs extraction (e.g., HR and RR), a classical PPG-extraction framework known as Spatial Redundancy [15], as illustrated in Fig. 3, is utilized in conjunction with living-skin detection [16] to find the pulsatile ROI. To isolate the image regions containing respiration and pulse, we divide the video frames into multi-scale blocks with different sizes. Compared to pulse, the ROI of motion-based RR tends to be a larger area encompassing the entire torso. Thus we segment the image into four scales:  $48 \times 48$ ,  $72 \times 72$ ,  $96 \times 96$ , and  $120 \times 120$  pixels for RR extraction, and utilize segmentation at four smaller scales:  $6 \times 6$ ,  $12 \times 12$ ,  $24 \times 24$ , and  $48 \times 48$  pixels for HR extraction, with a stride half of the block size per scale. To attain high efficiency of vital signs extraction, we selectively process the half image with upper body for PPG signal extraction, while employing the full image for motion signal extraction. This is reasonable since a person's head orientation towards the headboard or footboard remains relatively stable over the sleep period.

Second, there is minimal continuous body movement during sleep, and the most frequent motion is body rollover (sleep posture transition). Therefore, we only extract the PPG signal when the body remains still and ignore the period with occasional posture changes. The single IR-channel based PPG trace  $C_i(t)$  is created by averaging the spatial pixels within the  $i_{th}$  image block and then concatenating them along the temporal dimension. To eliminate the dependency on the illumination intensity, we construct the DC-normalized PPG traces by:

$$\tilde{C}_i(t) = \frac{C_i(t)}{\mu(C_i(t))} - 1, \quad (2)$$

where  $\mu(\cdot)$  denotes the time-averaging operation. Meanwhile, the PixFlow algorithm [51] in (3) is employed to extract the global motion signals between the temporal image sequence with the stride of two frames, in both vertical and horizontal directions. In (3),  $I$  denotes the image pixels within same image blocks,  $\frac{\partial I}{\partial x}$  and  $\frac{\partial I}{\partial y}$  represent image spatial gradients in the X

and Y directions, while  $v_x$  and  $v_y$  denote motion velocities in respective directions and the integration of them is the motion signal we need:

$$\frac{\partial I}{\partial t} = - \begin{bmatrix} \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix}. \quad (3)$$

The motion signal contains different motion sources in a video, including subtle respiratory motions and intense body rollover. The latter often involves global pixel movements, allowing the detection of body rollovers and other intense body movements by analyzing the motion statistically across all image blocks.

A more complex process is necessary for extracting subtle physiological signals. The PPG signal and motion signal are both divided into multiple short sliding windows for generating their power spectra using Fourier transform. Subsequently, to identify image blocks with high-quality candidates of PPG signals and motion signals, the signal-to-noise ratio (SNR) of each image block is calculated to render heatmaps of SNR. Based on the ranking of SNR, we select the top 10 image blocks from both the PPG heatmap and motion heatmap, used as the RoIs for PPG-based and motion-based RR extraction, respectively. Signals of selected ROI are combined into the final PPG signal and breath signal. For the reconstructed breath signal, a band-pass filter ([0.15, 1.0] Hz) is applied to remove intensive body motion, obtaining motion-based RR. Similarly, for the reconstructed PPG signal, band-pass filters with different frequency ranges (RR: [0.15, 1.0] Hz; HR: [0.6, 2.0] Hz) are applied to provide PPG-based RR and HR, respectively.

### C. Sleep Posture

It is challenging to analyze sleep posture in blurry images. To achieve this goal, we employ ResNet-18 [52] as the second branch of the proposed framework, and the fully connected layer's output is adjusted to classify images into five posture categories (off-bed/empty, supine, left-side, right-side, prone). However, the similar scenes and subjects' prolonged immobility during sleep result in high redundancy of collected frames in the training dataset, which leads to overfitting when using such a dataset to train the model. To alleviate this problem, first, a dropout probability of 0.2 was integrated to bolster the architecture's resilience to overfitting by randomly deactivating neurons during training, promoting the generalized feature learning for sleep posture classification. Second, using the Adam optimizer, with a learning rate of 0.0001, to fine-tune the network's weights by iteratively adjusting them in the direction of steeper gradients, facilitating faster convergence. Additionally, the weight decay coefficient ( $l_2$  regularization coefficient) of 0.001 encourages the model to minimize the magnitude of weights, preventing large weights from dominating the learning process and avoiding overfitting [53]. Another problem is the slow learning due to the minor differences among sleep postures, we use the cross-entropy loss function in (4) to guide the optimization of the network over the training of 60 epochs:

$$L_c(y_i, \hat{y}_i) = - \sum_{i=1}^c \hat{y}_i \log(y_i) \text{ with } y_i = \frac{\exp(w_i x)}{\sum_j^c \exp(w_j x)}, \quad (4)$$

where  $w_i$  denotes the  $i$ -th row vector of the weight matrix;  $x$  denotes the feature vector extracted from blurry images by ResNet-18;  $c$  is the number of sleep postures;  $y_i$  denotes the predicted sleep posture; and  $\hat{y}_i$  denotes the label.

## IV. EXPERIMENTAL SETUP

This section introduces the experimental setup, benchmark dataset, and evaluation metrics for assessing the sleep parameters measured by a defocused camera.

### A. Experimental Setup

In the dark chamber, the defocused monochrome IR camera (IDS UI-3860CP-M-GL R2 with Sony's 2.1 MP sensor IMX290) and an IR light source were positioned above the bed to capture the top-view infrared videos (see Fig. 2). To ensure sufficient image brightness and mimic the night monitoring condition, the light source emitted invisible light at 850 nm. An 850 nm narrow-band IR-pass filter was equipped inside the camera, exclusively transmitting infrared light to the image sensor. To mimic real sleep scenarios, volunteers were instructed to sleep in five common postures, following the order as shown in Fig. 4 (i.e., off-bed/empty, supine, left side, right side, and prone), with each posture remained for 2.5 minutes. Volunteers were allowed to change sleep postures freely according to their individual sleep habits. Besides, we used a bedsheets to cover the subject's body with the face exposed to mimic the challenge of occlusion in real sleep scenarios. The experiment involved 19 healthy volunteers (11 males and 8 females) with an average age of  $22.7 \pm 2.08$  years, average height of  $168.0 \pm 7.82$  cm, and average weight of  $57.6 \pm 10.38$  kg. The study was approved by the Institutional Review Board of Southern University of Science and Technology (No. 2023131), and written informed consent were obtained from the test subjects.

### B. Benchmark Dataset

**1) Video Dataset:** All videos were recorded in an uncompresssed data format, with a resolution of  $548 \times 968$  pixels at 20 frames per second. To validate the feasibility of using defocused cameras for sleep parameters measurement and to explore whether defocusing adversely affects the measurement, we recorded abundant videos containing five aforementioned sleep postures in both the focused and defocused conditions. We also recorded videos with and without bedsheets covers to investigate the impact of occlusion. Additionally, to evaluate performance under various degrees of blurring, we applied Gaussian filters with different blurring strengths to process the focused videos to simulate the defocused image blur (Note: The simulation is only used to study the effect of blur degree, such simulation will not be used as there is still a risk of privacy invasion before the image is digitally blurred). To simulate blurring with different levels close to real defocused conditions for a reasonable comparison, we specifically employed convolution kernels with various sizes and a uniform standard deviation of 25 pixels. Three settings were considered: *Simulated-1* with a (25, 25)-pixel kernel, *Simulated-2* with a (35, 35)-pixel kernel,

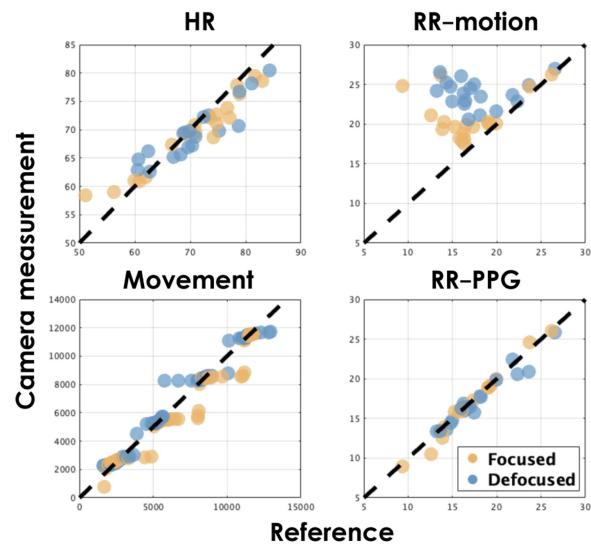
and *Simulated-3* with a (45, 45)-pixel kernel. In conclusion, a total of five video conditions were experimented under covered and uncovered conditions, and finally, each subject had 10 videos recorded with a combination of different protocols as shown in Fig 4.

The pretraining on the public dataset of ImageNet enables ResNet-18 to extract basic image features. For sleep posture classification, videos from all categories of 15 participants were employed as the training set to retrain the model, while videos from all categories of 4 participants served as the test set. Both the training and test datasets were constructed by uniformly sampling the videos of each volunteer and sleep posture. During the period of the same sleep posture, many frames are visually similar. To increase the data scarcity, we applied data augmentation to the original image set, introducing variations in noise levels, random occlusions, brightness ranges, blurring, and other challenge factors. For the test set, to investigate the influence of different factors on sleep posture recognition, five different subsets of the test set were generated: *focused* (only focused images), *defocused* (only defocused images), *covered* (only covered images), *uncovered* (only uncovered images), and *overall* (images concerns all situations). Each subset comprises an equal number of images in five sleep postures. Additionally, to quantify the impact of blurring on sleep posture recognition, including both focused and defocused situations, coupled with simulations based on the Gaussian filter, the subsets of blurred images with 11 different blurring levels were generated. In conclusion, we have obtained a mixed training set comprising 3750 images, as well as 16 different test sets with each containing 1000 images.

**2) Ground-Truth:** To benchmark the camera-based contactless sleep parameters measurement, the reference of physiological signals was recorded by Benevision N17 patient monitor (Mindray, China), which has three-lead ECG and PPG. By applying band-pass filters to the ECG signal across different frequency ranges (HR: [0.6, 2.0] Hz; RR: [0.15, 1.0] Hz; Movement: [0.0, 0.1] Hz), we created the ground truths for HR, RR, and movement. The ground truth along with their corresponding timestamps were saved and then exported for evaluation. For the labels of sleep posture, since all subjects strictly follow the order of posture changes specified in our protocol, we used the protocol to generate the posture labels.

### C. Evaluation Metrics

**1) Heart Rate, Respiration Rate and Movement:** To evaluate the performance of the proposed method, we utilized the mean absolute error (MAE) to assess the HR and RR measurements from 19 participants across various situations (including five sleep postures, with or without bedsheets occlusion, and settings of focused, defocused, and simulated blurring). For the evaluation of HR and RR, given the presence of posture changes during recording (necessary for validating the sleep parameters measurement in different sleep postures), significant body movements need to be addressed. Therefore, we employed the intensity of motion signals as a metric to exclude signal segments with large movements to improve the accuracy of statistical evaluation. This motion metric is derived by



**Fig. 6.** For HR and RR, the correlation between camera measurements and references, where each scatter represents a single subject. For movement, each scatter represents the posture transition moment of the camera and reference, and each subject has four scatters per recording. The dotted lines denote regression with  $R = 1$ .

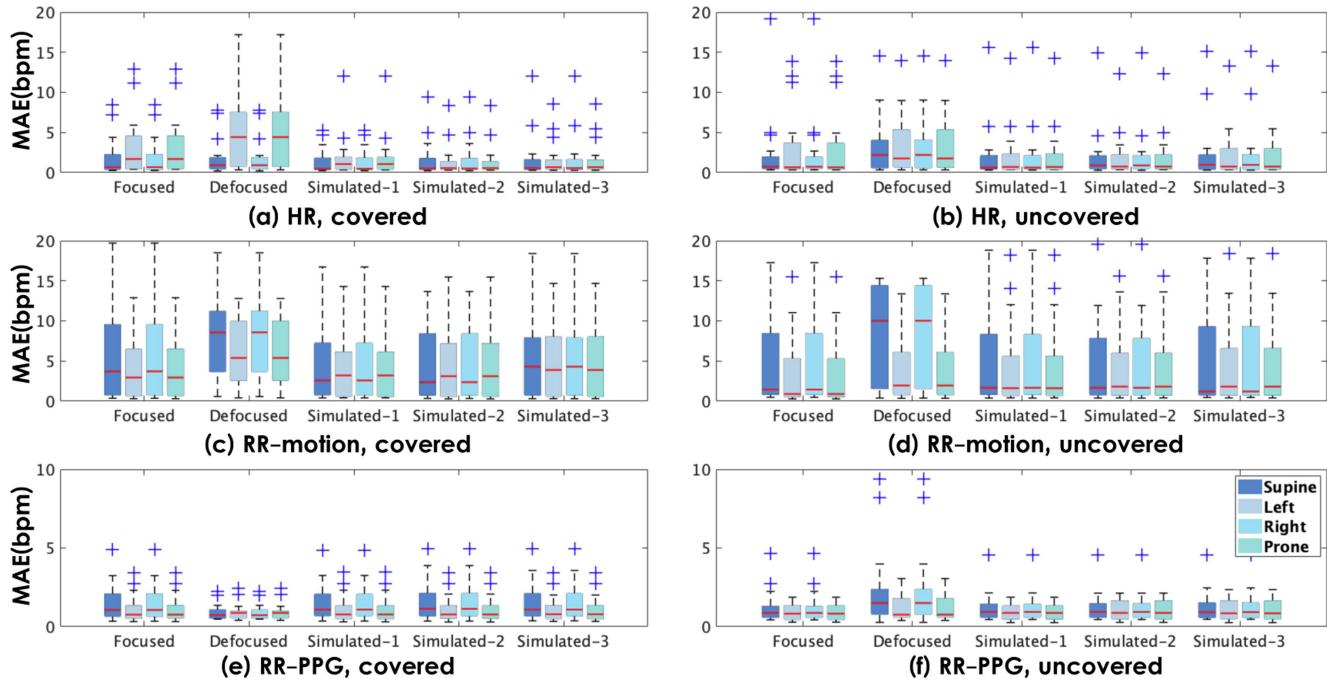
computing the temporal standard deviation of motion signals. The subtle movements (including respiratory motion) during sleep typically have a motion intensity smaller than 1 in our method, i.e., threshold = 1 is used to exclude large motions that are considered not relevant to HR and RR, resulting in a more accurate assessment. But for the movement signal, we use a sliding window to find the peaks in the area of intensity greater than 1, which represents the time slot with strong movement, i.e., typically when rollover occurs. To compare the measurements of the camera and patient monitor on the same scale, we normalized the movement signals from both the camera and patient monitor using min-max normalization. Since the experiment includes 5 sleep postures, there are 4 posture transitions. The duration and amplitude of sleep posture transitions varied among individuals. We used the timestamp of maximum amplitude as the marker for a transition, i.e., the peaks of the movement waveform represent the moments of rollover occurrences. The K-means clustering was utilized to find the time centers of the camera and reference, pinpointing the overall moment when rollover starts. Finally, the movement detection was assessed by calculating the time offset of each time center between the camera and reference.

The Pearson correlation coefficient is utilized to analyze the correlation between camera measurements of the three aforementioned sleep parameters and the reference. We use p-values to indicate the significance of correlations.

**2) Sleep Posture:** We employ the standard machine learning metrics to evaluate the classification of five postures, including accuracy, precision, recall, and F1-score.

## V. RESULTS AND DISCUSSION

In this section, we first discuss the evaluation results of privacy protection, and then the results of HR, RR, and movement. The results for sleep posture detection are discussed separately. Finally, a summary is given.



**Fig. 7.** The MAE of HR and RR between camera and reference under multiple conditions. As the first sleep posture is off-bed (no subject), boxplots for all types of videos only show the MAE for four in-bed sleep postures.



**Fig. 8.** The first row is recorded by the defocused camera, and the second row is deblurred using the state-of-the-art method.

#### A. Facial Privacy Protection Test

To illustrate the influence of defocus blurring on the de-identification, we validate it using advanced facial analysis tools, and the snapshots are exemplified in Fig. 5. The three tasks (face detection, face landmark detection, and face recognition) based on the hybrid face analysis framework in [54], are only functional in the focused condition. In the case of face recognition, we also found an increase in cosine similarity [55] with the increase of blur, which indicates that the disparities between the two images are progressively increasing. These results show that both the real and simulated defocus blur can successfully disrupt face analysis tasks, highlighting the effectiveness of protecting privacy. We also see that as the degree of blur increases, it leads to more secured identity protection. In the image restoration using a state-of-the-art technique [56], it cannot recover the optically blurred face recorded by our defocused camera (see Fig. 8). This shows the promising performance of a simple adjustment of the camera focal length for protecting the privacy.

#### B. Heart Rate, Respiration Rate and Movement

Table II presents the MAE of three indices extracted from videos of all categories from 19 volunteers: HR and RR based on PPG, and motion-based RR, in comparison with the reference from the patient monitor. For HR measurements, the overall MAE is slightly lower in three simulated situations compared to the focused condition, while the worst performance is observed in the defocused condition (see Fig. 7(a) and (b) and Table II). The comparison between the focused and defocused modes, as well as the performance degradation evident in three simulations, is foreseeable since image blurring combines skin pixels and non-skin pixels, which pollutes the PPG measurement. However, the enhancement of three simulations compared to the focused and defocused modes can be attributed to the Gaussian filtering of focused videos. The post-filtering using a Gaussian kernel somehow reduces the image sensor noise and improves the image SNR, thereby resulting in better performance of PPG measurement.

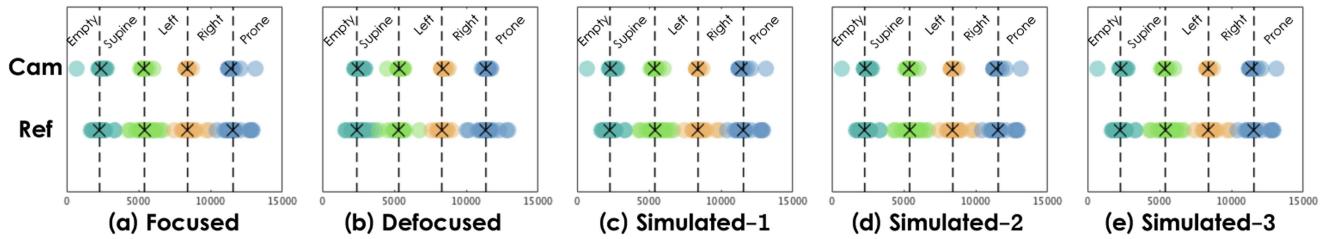
To better understand the impact of blurring, we compared three simulated situations and found that blurring worsens HR measurement, as indicated by the overall MAE differences between focused and defocused situations. Normally, the MAE of a reliable measurement of HR and RR is required to be smaller than 5 bpm w.r.t. the reference [57]. Most HR measurements under defocused conditions (see Table II) also fluctuate around that standard. Meanwhile, Fig. 6 demonstrates a high correlation (HR:  $R = 0.9076$ ,  $p < 0.000$ ) between camera measurements and the reference, further confirming the effectiveness of using a defocused camera for HR monitoring. Given that facial skin is typically exposed to the defocused camera, we can extract PPG

**TABLE II**  
MAE OF HR AND RR MONITORING IN EACH CONDITION

Subject	Heart Rate MAE±SD (bpm)					Respiratory Rate-motion MAE±SD (bpm)					Respiratory Rate-PPG MAE±SD (bpm)				
	Focused	Defocused	Simulated-1	Simulated-2	Simulated-3	Focused	Defocused	Simulated-1	Simulated-2	Simulated-3	Focused	Defocused	Simulated-1	Simulated-2	Simulated-3
1	3.2±3.2	2.8±3.6	2.5±2.8	2.2±3.1	2.0±2.7	12.7±11.3	8.1±4.6	12.9±12.1	12.3±11.8	11.3±10.6	0.9±0.9	1.3±1.3	1.0±0.9	1.0±0.9	1.0±1.0
2	1.4±2.1	5.0±4.6	1.6±2.7	1.1±1.7	1.4±2.5	2.1±2.5	7.8±5.0	1.7±2.6	2.2±4.0	2.9±4.3	1.0±1.1	1.2±1.1	0.9±1.0	0.9±1.0	0.9±1.0
3	3.5±4.0	5.8±4.2	2.5±3.8	2.6±3.8	3.9±4.9	1.6±2.5	2.4±2.0	1.7±2.6	1.9±2.7	2.3±3.0	1.0±1.3	3.2±3.4	1.0±1.1	1.1±1.4	1.1±1.5
4	1.8±2.4	7.9±8.1	1.0±1.8	0.6±0.8	1.1±1.8	0.6±0.6	1.3±1.6	0.5±0.6	0.6±0.6	0.6±0.6	0.5±0.6	1.4±2.3	0.6±0.6	0.6±0.6	0.6±0.6
5	7.2±8.3	4.4±3.2	3.5±5.3	2.4±3.9	2.4±3.6	1.2±1.4	2.1±1.8	1.4±1.6	1.3±1.6	1.3±1.5	1.0±1.2	0.5±0.5	1.0±1.1	1.0±1.0	0.9±1.0
6	3.3±4.0	10.5±9.0	2.3±2.9	2.2±3.2	1.9±2.8	1.5±1.7	11.5±6	1.5±1.7	1.5±1.7	1.5±1.7	1.5±1.7	1.9±1.6	1.5±1.8	1.5±1.7	1.5±1.7
7	2.2±4.1	7.1±6.6	1.6±3.2	1.6±3.3	1.5±2.7	6.3±5.1	3.5±3.7	5.9±4.6	6.5±4.5	7.0±5.0	1.5±1.4	1.0±1.2	1.5±1.4	1.5±1.4	1.4±1.4
8	5.7±7.4	7.6±7.2	1.5±2.9	2.3±4.2	2.0±3.4	3.9±3.5	6.3±4.3	2.8±2.8	3.1±3.0	3.8±3.0	1.0±1.0	1.0±1.2	1.0±1.0	1.0±1.0	1.0±1.0
9	4.0±3.6	2.2±2.5	3.5±3.2	3.9±3.4	4.0±3.6	0.9±1.0	11.0±5.9	0.9±0.9	0.9±1.0	0.9±0.9	1.1±1.4	1.4±1.2	1.0±1.1	1.2±1.5	1.3±1.6
10	3.3±2.9	4.7±4.9	3.2±3.8	3.0±3.6	4.9±4.7	1.5±2.0	7.8±5.6	1.1±1.2	1.0±1.1	1.1±1.4	1.0±1.0	2.5±2.6	1.1±1.2	1.1±1.2	1.1±1.3
11	5.0±6.9	4.0±2.6	6.5±4.9	7.3±5.7	7.6±6.0	6.2±4.1	13.0±5.6	4.6±3.9	4.7±4.0	4.6±4.1	3.0±1.6	1.2±1.1	2.8±1.5	2.9±1.5	2.8±1.4
12	8.3±8.1	5.2±5.3	4.1±5.4	4.5±5.1	3.3±3.5	2.7±2.8	5.5±4.8	2.0±2.9	1.9±2.7	2.2±2.8	0.9±1.1	1.4±1.8	0.9±1.1	1.1±1.2	1.0±1.2
13	2.8±2.9	2.5±3.8	1.7±2.8	1.9±3.1	1.9±3.1	1.6±1.8	2.9±2.8	1.7±1.9	1.6±1.8	1.4±1.4	2.1±1.2	2.4±2.2	1.3±1.2	1.3±1.3	1.3±1.3
14	3.6±2.8	6.8±6.2	2.3±2.4	1.6±2.1	1.9±2.6	2.8±3.7	7.7±4.6	2.7±3.7	3.1±4.1	2.9±3.9	1.2±1.2	1.2±1.5	1.2±1.2	1.3±1.2	1.5±1.6
15	4.7±4.9	5.1±4.1	2.0±3.1	1.9±3.0	1.1±2.1	3.3±3.3	1.0±1.4	4.6±4.6	5.6±5.3	5.4±5.1	1.3±1.2	2.2±2.2	1.3±1.2	1.4±1.3	1.4±1.3
16	9.8±8.0	4.6±5.3	7.4±8.5	7.3±8.9	7.7±9.6	15.4±5.9	6.8±4.7	15.9±6.1	16.6±6.0	16.6±6.4	1.5±1.4	1.3±1.3	1.5±1.4	1.5±1.4	1.5±1.4
17	6.9±6.0	3.7±4.4	6.4±5.9	6.3±5.4	6.1±6.3	8.6±5.6	4.8±5.0	9.5±6.4	10.6±7.1	11.8±8.9	2.7±1.7	0.9±1.1	2.7±1.6	2.7±1.8	2.7±1.7
18	4.4±5.2	4.1±5.7	3.3±4.4	3.3±5.6	3.1±5.0	4.8±4.5	10.0±6.9	6.5±4.7	8.4±4.9	8.6±4.4	1.4±1.5	1.2±1.3	1.3±1.5	1.3±1.4	1.3±1.4
19	4.5±3.1	3.9±4.6	3.4±2.6	3.9±4.3	4.5±6.3	2.1±3.0	10.1±5.8	3.8±6.8	3.9±5.0	3.9±4.8	0.5±0.6	1.1±1.0	0.5±0.6	0.5±0.6	0.5±0.6
Overall	4.5±4.7	5.2±4.9	3.2±3.8	3.2±3.9	3.3±4.1	4.2±3.5	6.5±4.3	4.3±3.8	4.6±3.8	4.7±3.9	1.3±1.2	1.5±1.6	1.3±1.2	1.3±1.2	1.3±1.3

\*“Focused” and “Defocused” denote the corresponding camera condition, and “Simulated-X” denotes the simulated blurry conditions using a Gaussian filter.

\* MAE: mean absolute error; SD: standard deviation.



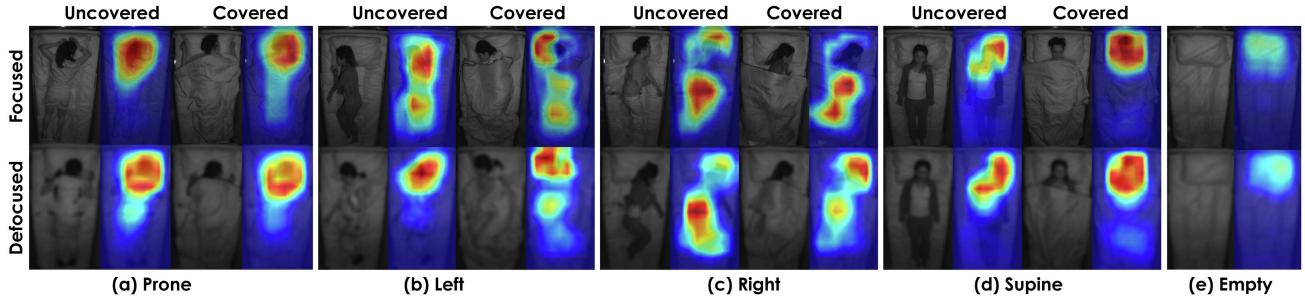
**Fig. 9.** Comparison of movement between camera and reference. **Cam** refers to camera measurements, while **Ref** refers to the reference from the patient monitor. The four different colors represent time points of posture transition for all subjects during sleep. The crosses denote the cluster centers of posture transition measured by the camera and reference.

signals from the face without facial details, regardless of whether the body is covered or not. Consequently, the results shown in Fig. 7(a) and (b) are reasonable, indicating that occlusion has no significant impact on HR measurement. The poorer performance in the left-side and prone sleep postures might be attributed to the reduced skin regions for PPG measurement.

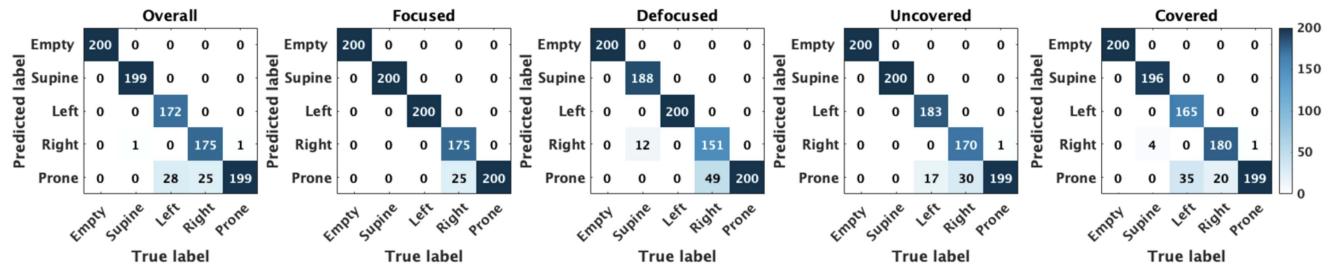
For two different types of RR measurements: motion-induced RR (RR-motion) and blood pulsation-based RR (RR-PPG) are used for benchmarking. From Table II and Fig. 7, it is clear that the overall MAE of RR-PPG is noticeably smaller than that of RR-motion, which suggests that RR-PPG is more robust when camera is in the defocused mode. As shown in Fig. 6, in the defocused situation, RR-PPG shows a much stronger correlation with the reference ( $R = 0.9734$ ,  $p < 0.000$ ) than RR-motion ( $R = 0.0566$ ,  $p = 0.8179$ ), and the latter is likely due to negative impact of image blurring on pixel motion analysis, i.e., blurring is particularly harmful for subtle motion estimation. Besides, for RR-motion specifically, Fig. 7(c) and (d) illustrate that occlusion also worsens its measurements as it can hinder the measurement of pixel movements from body areas. The poorer performance of supine posture is because that respiratory motion is perpendicular to the camera viewing angle, thus the respiratory motion is diminished in the video. Compared with the left side, the worse performance of the right side is probably due to other factors, such as uneven illumination caused by wall reflections and constraints imposed by the wires of the monitor

while breathing in a sleep posture facing away the monitor. For RR-PPG (see Fig. 7(e) and (f)), it performs exceptionally well across all conditions, immune to occlusions and sleep postures. This further illustrates the robustness of RR-PPG under the defocused situation. In the sleep case, the absence of voluntary movements and head support of the pillow allows the defocused camera to better measure the respiratory modulation from facial PPG signals. However, there is a quality drop with RR-PPG due to image blurring, the reason for which is similar to the degradation of HR.

Fig. 9 shows the results of movement measurement in all defocused conditions, where the camera can accurately detect four posture transitions. The four reference clusters show more scattering in the X direction compared to camera measurements, as the monitor employs contact-based measurements, introducing additional interference from the motion of cables. Table III provides the time offsets of each transition between camera measurement and the reference in all defocused conditions. It can be observed that blurring causes a minimal time delay (approximately 2 seconds) in camera measurement, and it shows a strong correlation with the reference (see Fig. 6,  $R = 0.9946$ ,  $p < 0.000$ ) in the defocused situation. Besides, the adjustable threshold enables the camera to accurately capture a wider range of movements with different amplitudes. However, Table III indicates that blurring leads to a slightly increased temporal offset between camera measurements and reference, which is



**Fig. 10.** Feature heatmap of sleep posture classification, where deeper red indicates that the model assigns higher weights to the corresponding image region.



**Fig. 11.** Confusion matrices for sleep posture recognition under various recording conditions. In each condition, there is an equal number of images across all sleep postures.

**TABLE III**  
TIME OFFSET OF POSTURE TRANSITION BETWEEN CAMERA MEASUREMENT  
AND REFERENCE UNDER DIFFERENT CONDITIONS

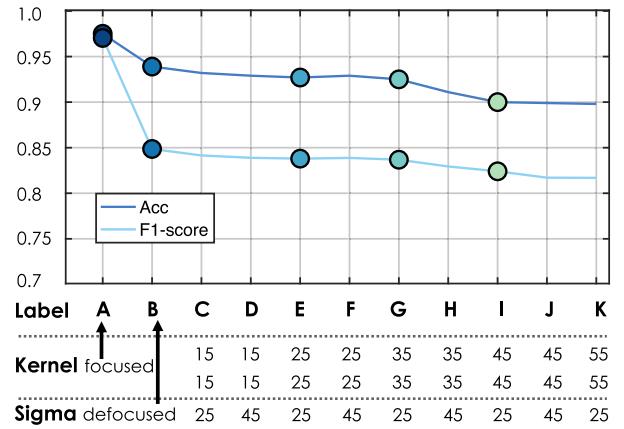
Condition (sec)	E→S	S→L	L→R	R→P	Average
Focused	4.7	-2.9	-2.2	-6.1	-1.6
Defocused	3.6	1.5	3.1	0.5	2.2
Simulated-1	3.8	-3.1	-2.0	-6.1	-1.8
Simulated-2	3.9	-3.1	-2.1	-6.1	-1.9
Simulated-3	3.9	-3.1	-2.2	-6.1	-1.9

\*\*“E”, “S”, “L”, “R”, and “P” refers to five sleep postures, namely empty, supine, left-side, right-side, and prone.

\* →: sleep posture transition.

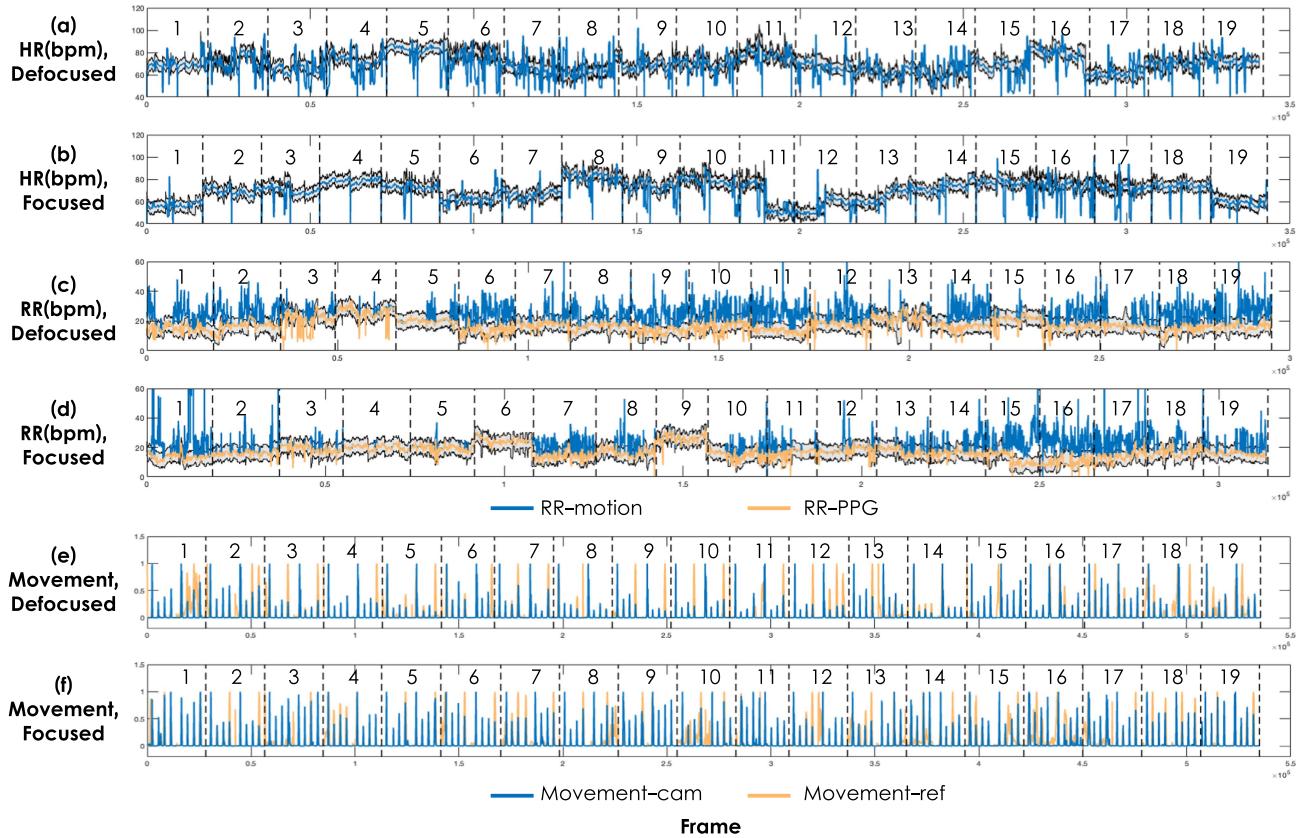
similar to the phenomenon of performance degradation in the measurement of other sleep parameters.

Fig. 13 shows the long-term traces of HR, two types of RR, and movement under focused and defocused conditions. For HR measurement, there is an increase in noise under defocused conditions. In terms of RR-motion measurement, both conditions have low SNR caused by insufficient illumination in the chest and abdomen, with noise exacerbated in the defocused situation; however, RR-PPG shows robust performance in both cases. For movement measurement, accurate occurrence time of rollover is detected in both conditions (peaks of camera measurement align with the reference closely), yet the movement signal of the camera is weaker (it shows a smaller amplitude) in defocused conditions, because blurring makes the camera less sensitive to the pixel-level motion, posing a challenge in distinguishing different subtle movements. Fig. 14 shows the coverage for HR and two types of RR under both the focused and defocused

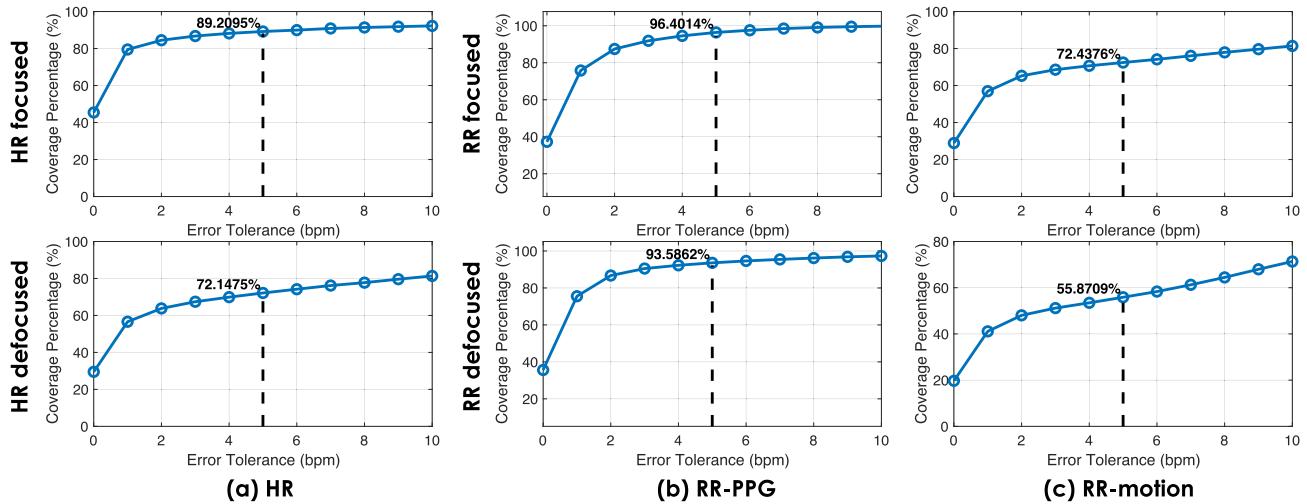


**Fig. 12.** The relationship between the performance of sleep posture recognition and blurring level. Apart from the recorded videos in the focused and defocused modes, simulations were achieved using Gaussian filters, with corresponding convolution kernel sizes and standard deviations shown at the bottom. The circles represent five levels of blurriness: Focused, defocused, simulated-1, simulated-2, and simulated-3.

conditions. For acceptable HR estimation (reference  $\pm 5$  bpm), we achieved a measurement coverage of 72.1% in the defocused condition, which is approximately 17.1% lower than the focused mode. It is attributed to real defocus, which mixes non-skin pixels, sensor noise, and useful skin-pixels together, reducing the SNR of the PPG signal. Regarding RR, the RR-motion performance is inferior to RR-PPG performance, with the latter achieving a measurement coverage of 93.6% even in the



**Fig. 13.** Time-domain waveforms of HR, RR, and movement for 19 subjects under both focused and defocused conditions. For HR and RR waveforms, the gray region represents a range of  $\pm 5$  around the monitor's reference. For movement, the suffix “cam” denotes camera measurement, while the suffix “ref” denotes reference.



**Fig. 14.** Under focused and defocused situations, the coverage of HR and RR estimation over the entire period at various error tolerances. The dashed line indicates the acceptable tolerance of error ( $\pm 5$  bpm).

defocused condition. These results demonstrate the feasibility of using a defocused camera for HR and RR extraction, but further improvement is needed for HR measurement, especially improving the PPG signal extraction from optically blurred images.

### C. Sleep Posture

As shown in Table IV, in a mixed test set including defocus and occlusion, the sleep posture recognition achieved an overall accuracy of 94.5%. Both occlusion and defocus can

TABLE IV

METRICS OF SLEEP POSTURE RECOGNITION UNDER VARIOUS VIDEO CONDITIONS

Condition	Posture	Precision	Recall	F1-score	Accuracy
Overall	Empty	1.0	1.0	1.0	0.945
	Supine	1.0	0.995	0.997	
	Left	1.0	0.860	0.925	
	Right	0.989	0.875	0.929	
	Prone	0.790	0.995	0.881	
Focused	Empty	1.0	1.0	1.0	0.975
	Supine	1.0	1.0	1.0	
	Left	1.0	1.0	1.0	
	Right	1.0	0.875	0.933	
	Prone	0.889	1.0	0.941	
Defocused	Empty	1.0	1.0	1.0	0.939
	Supine	1.0	0.940	0.969	
	Left	1.0	1.0	1.0	
	Right	0.926	0.755	0.832	
	Prone	0.803	1.0	0.891	
Uncovered	Empty	1.0	1.0	1.0	0.952
	Supine	1.0	1.0	1.0	
	Left	1.0	0.915	0.956	
	Right	0.994	0.850	0.916	
	Prone	0.809	0.995	0.892	
Covered	Empty	1.0	1.0	1.0	0.940
	Supine	1.0	0.980	0.990	
	Left	1.0	0.825	0.904	
	Right	0.973	0.900	0.935	
	Prone	0.783	0.995	0.876	

hinder sleep posture recognition, and defocus leads to worse performance. However, we still achieved an accuracy of 93.9% on a test set containing only defocused images. The heatmap for the sleep posture classification is shown in Fig. 10, presenting intermediate results (attention of network) for five sleep postures under focused and defocused situations, as well as uncovered and covered situations. When lying on the left or right side, the model focuses on the entire torso, especially the chest and hips with more shadows. However, when lying supine or prone, it focuses on the facial features where the most distinct texture differences are presented. The model shows a more uniform distribution of features when recognizing the empty state compared to other postures. The weight distribution of different sleep postures further confirms that blurry images captured by the defocused camera can be used for sleep posture classification. From other metrics and the confusion matrices for various conditions in Fig. 11, it can be observed that both defocus and occlusion can cause other sleep postures to be incorrectly classified as prone. This challenge primarily arises from the irregular orientation of the head towards the left or right when sleeping in a prone posture. Apart from the video sets already tested for other sleep parameters, Fig. 12 shows a finer scale assessment of ResNet-18 based model, which indicates that it maintains good performance across a wide range of blur levels, as the accuracy is still up to 80% under the most blurry conditions.

In summary, defocus successfully disrupted three face analysis tasks, demonstrating that defocus can effectively prevent the

capture of facial details. Both the PPG-based HR and RR estimation, as well as movement, have strong correlations with reference. For sleep posture recognition, the ResNet-18 based model achieves an overall accuracy of 94.5%. All the results show that the defocused camera is a feasible and promising option for a comprehensive sleep parameters measurement. Although there are some negative influences of occlusion on RR-motion and sleep posture recognition, along with the slight impact of different sleep postures on HR measurement, they are expected to be solved from algorithmic aspects in the future. In our study, optical blurring itself indeed has a negative impact on all monitored sleep parameters. To strike a good trade-off of privacy protection versus the quality of measurement, an appropriate level of optical blur is needed in practical applications. We propose to further explore how to automatically control the degree of defocus for optimal measurement of sleep parameters while still protecting the privacy. No complex hardware modifications nor cumbersome image encryption algorithms are needed, except for a lens that permits focal length adjustment. For the next step, we will explore the integration of radar or camera-based seismocardiogram [58] that are privacy-insensitive monitoring techniques for capturing signals from other body parts, to enhance robustness in the presence of occlusion and conduct studies in real-world application scenarios, including sleep centers and senior centers, to further investigate privacy-protected sleep health monitoring. Additional vital signs that may facilitate sleep monitoring (e.g., HRV, SpO<sub>2</sub>, sleep staging) will be explored as well in the fusion with other devices such as radar [59], thermal camera [60], microphone [61].

## VI. CONCLUSION

To solve the issue of privacy invasion of camera-based sleep monitoring, we propose to use a defocused camera without additional hardware modifications except for a simple adjustment of the camera focal length. We employed a spatial redundancy framework in conjunction with living-skin detection to process blurry images captured by a defocused camera, extracting HR, RR, and movement with a reasonably good performance. Additionally, we utilized ResNet-18 to extract features of body posture from blurry images, achieving robust recognition of five common sleep postures. The experiments show that our approach is robust to occlusions like bedsheets covers. Furthermore, by selecting an appropriate level of optical blurring, it is possible to achieve reliable measurement of the aforementioned four sleep parameters while protecting the privacy of the monitored subject. For future work, we aim to conduct studies in real-world application scenarios (e.g. hospitals, sleep centers and senior centers) and enrich the sleep parameters that can be measured by a remote defocused camera.

## REFERENCES

- [1] A. A. Bucklin et al., "High prevalence of sleep-disordered breathing in the intensive care unit—a cross-sectional study," *Sleep Breathing*, vol. 27, no. 3, pp. 1013–1026, 2023.

- [2] K. Watanabe, Y. Kurihara, and H. Tanaka, "Ubiquitous health monitoring at home—sensing of human biosignals on flooring, on tatami mat, in the bathtub, and in the lavatory," *IEEE Sensors J.*, vol. 9, no. 12, pp. 1847–1855, Dec. 2009.
- [3] D. Goyal, C. Gay, and K. Lee, "Fragmented maternal sleep is more strongly correlated with depressive symptoms than infant temperament at three months postpartum," *Arch. Women's Ment. Health*, vol. 12, pp. 229–237, 2009.
- [4] Z. Hussain, Q. Z. Sheng, W. E. Zhang, J. Ortiz, and S. Pouriyeh, "Non-invasive techniques for monitoring different aspects of sleep: A comprehensive review," *ACM Trans. Comput. Healthcare*, vol. 3, no. 2, pp. 1–26, 2022.
- [5] H. P. Chang, Y. F. Chen, and J. K. Du, "Obstructive sleep apnea treatment in adults," *Kaohsiung J. Med. Sci.*, vol. 36, no. 1, pp. 7–12, 2020.
- [6] Y. Lee et al., "A novel non-contact heart rate monitor using impulse-radio ultra-wideband (IR-UWB) radar technology," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 13053.
- [7] M. Harford, J. Catherall, S. Gerry, J. D. Young, and P. Watkinson, "Availability and performance of image-based, non-contact methods of monitoring heart rate, blood pressure, respiratory rate, and oxygen saturation: A systematic review," *Physiol. Meas.*, vol. 40, no. 6, 2019, Art. no. 06TR01.
- [8] S. M. Mohammadi, S. Enshaeifar, A. Hilton, D.-J. Dijk, and K. Wells, "Transfer learning for clinical sleep pose detection using a single 2D IR camera," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 290–299, 2021.
- [9] B. C. Ko, "A brief review of facial emotion recognition based on visual information," *Sensors*, vol. 18, no. 2, 2018, Art. no. 401.
- [10] L. Niu, J. Speth, N. Vance, B. Sporrer, A. Czajka, and P. Flynn, "Full-body cardiovascular sensing with remote photoplethysmography," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 5994–6004.
- [11] P. Voigt and A. V. d. Bussche, *The EU General Data Protection Regulation (GDPR), A Practical Guide*, 1st ed., vol. 10. Cham, Switzerland: Springer, 2017, pp. 10–5555.
- [12] S. L. Pardau, "The California consumer privacy act: Towards a European-style privacy regime in the United States," *J. Tech. L. Pol'y*, vol. 23, pp. 68–114, 2018.
- [13] A. A. Chaaraoui, J. R. Padilla-López, F. J. Ferrández-Pastor, M. Nieto-Hidalgo, and F. Flórez-Revuelta, "A vision-based system for intelligent monitoring: Human behaviour analysis and privacy by context," *Sensors*, vol. 14, no. 5, pp. 8895–8925, 2014.
- [14] W. Wang, A. C. d. Brinker, and G. d. Haan, "Single-element remote-PPG," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 7, pp. 2032–2043, Jul. 2019.
- [15] W. Wang, S. Stuijk, and G. D. Haan, "Exploiting spatial redundancy of image sensor for motion robust rPPG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 415–425, Feb. 2015.
- [16] W. Wang, S. Stuijk, and G. d. Haan, "Living-skin classification via remote-PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2781–2792, Dec. 2017.
- [17] V. P. Tran, A. A. Al-Jumaily, and S. M. S. Islam, "Doppler radar-based non-contact health monitoring for obstructive sleep apnea diagnosis: A comprehensive review," *Big Data Cogn. Comput.*, vol. 3, no. 1, pp. 3–23, 2019.
- [18] L. Ren, L. Kong, F. Foroughian, H. Wang, P. Theilmann, and A. E. Fathy, "Comparison study of noncontact vital signs detection using a doppler stepped-frequency continuous-wave radar and camera-based imaging photoplethysmography," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 9, pp. 3519–3529, Sep. 2017.
- [19] Y. S. Lee, P. N. Pathirana, T. Caelli, and S. Li, "Further applications of doppler radar for non-contact respiratory assessment," in *Proc. IEEE 35th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2013, pp. 3833–3836.
- [20] J. Xiong, H. Hong, H. Zhang, N. Wang, H. Chu, and X. Zhu, "Multitarget respiration detection with adaptive digital beamforming technique based on SIMO radar," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 11, pp. 4814–4824, Nov. 2020.
- [21] S. Yue, Y. Yang, H. Wang, H. Rahul, and D. Katabi, "BodyCompass: Monitoring sleep posture with wireless signals," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 2, pp. 1–25, 2020.
- [22] M. Schätz, A. Procházka, J. Kuchynka, and O. Vyšata, "Sleep apnea detection with polysomnography and depth sensors," *Sensors*, vol. 20, no. 5, 2020, Art. no. 1360.
- [23] F. King, R. Kelly, and C. G. Fletcher, "Evaluation of LiDAR-derived snow depth estimates from the iPhone 12 pro," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 7003905.
- [24] A. P. Addison, P. S. Addison, P. Smit, D. Jacquel, and U. R. Borg, "Non-contact respiratory monitoring using depth sensing cameras: A review of current literature," *Sensors*, vol. 21, no. 4, 2021, Art. no. 1135.
- [25] B. Deng, B. Xue, H. Hong, C. Fu, X. Zhu, and Z. Wang, "Decision tree based sleep stage estimation from nocturnal audio signals," in *Proc. IEEE 22nd Int. Conf. Digit. Signal Process.*, 2017, pp. 1–4.
- [26] S. Coşar, Z. Yan, F. Zhao, T. Lambrou, S. Yue, and N. Bellotto, "Thermal camera based physiological monitoring with an assistive robot," in *Proc. IEEE 40th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2018, pp. 5010–5013.
- [27] Z. Chen and Y. Wang, "Remote recognition of in-bed postures using a thermopile array sensor with machine learning," *IEEE Sensors J.*, vol. 21, no. 9, pp. 10428–10436, May 2021.
- [28] M. Hu et al., "Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and heart rates under sleep situation," *PLoS One*, vol. 13, no. 1, 2018, Art. no. e0190466.
- [29] W. Wang, S. Stuijk, and G. D. Haan, "A novel algorithm for remote photoplethysmography: Spatial subspace rotation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 9, pp. 1974–1984, Sep. 2016.
- [30] W. Wang and A. C. d. Brinker, "Algorithmic insights of camera-based respiratory motion extraction," *Physiol. Meas.*, vol. 43, no. 7, 2022, Art. no. 075004.
- [31] Y. Zhu, Y. Zeng, D. Huang, J. Huang, H. Lu, and W. Wang, "Occlusion-robust sleep posture detection using body rolling motion in a video," in *Proc. IEEE 45th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2023, pp. 1–5.
- [32] G. Liao, C. Shan, and W. Wang, "Comparison of PPG and BCG features for camera-based blood pressure estimation by ice water stimulation," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Inform.*, 2022, pp. 1–4.
- [33] Y. Ye, L. Pan, D. Yu, D. Gu, H. Lu, and W. Wang, "Notch RGB-camera based SpO<sub>2</sub> estimation: A clinical trial in a neonatal intensive care unit," *Biomed. Opt. Exp.*, vol. 15, no. 1, pp. 428–445, 2024.
- [34] D. Huang et al., "Generalized camera-based infant sleep-wake monitoring in NICUs: A multi-center clinical trial," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 5, pp. 3015–3028, 2024, doi: [10.1109/JBHI.2024.3371687](https://doi.org/10.1109/JBHI.2024.3371687).
- [35] H. Shu, L. Ren, L. Pan, D. Huang, H. Lu, and W. Wang, "Single image based infant body height and weight estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 6052–6059.
- [36] S. Ravi, P. Climent-érez, and F. Florez-Revuelta, "A review on visual privacy preservation techniques for active and assisted living," *Multimedia Tools Appl.*, vol. 83, no. 5, pp. 14715–14755, 2024.
- [37] K. Chung, K. Song, K. Shin, J. Sohn, S. H. Cho, and J. H. Chang, "Noncontact sleep study by multi-modal sensor fusion," *Sensors*, vol. 17, no. 7, 2017, Art. no. 1685.
- [38] M. v. Gastel, S. Stuijk, S. Overeem, J. P. v. Dijk, M. M. v. Gilst, and G. d. Haan, "Camera-based vital signs monitoring during sleep—a proof of concept study," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 5, pp. 1409–1418, May 2021.
- [39] Q. Wang, H. Cheng, and W. Wang, "Feasibility of exploiting physiological and motion features for camera-based sleep staging: A clinical study," in *Proc. IEEE 45th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2023, pp. 1–5.
- [40] N. A. Eiseman, M. B. Westover, J. M. Ellenbogen, and M. T. Bianchi, "The impact of body posture and sleep stages on sleep apnea severity in adults," *J. Clin. Sleep Med.*, vol. 8, no. 6, pp. 655–666, 2012.
- [41] P. K. Stein and Y. Pu, "Heart rate variability, sleep and sleep disorders," *Sleep Med. Rev.*, vol. 16, no. 1, pp. 47–66, 2012.
- [42] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, vol. 27, pp. 1–9.
- [43] X. Yang et al., "Towards face encryption by generating adversarial identity masks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 3897–3907.
- [44] P. Korshunov and T. Ebrahimi, "Using face morphing to protect privacy," in *Proc. IEEE 10th Int. Conf. Adv. Video Signal Based Surveill.*, 2013, pp. 208–213.
- [45] A. Tonge and C. Caragea, "Image privacy prediction using deep neural networks," *ACM Trans. Web*, vol. 14, no. 2, pp. 1–32, 2020.
- [46] A. Cui and S. J. Stolfo, "A quantitative analysis of the insecurity of embedded network devices: Results of a wide-area scan," in *Proc. 26th Annu. Comput. Secur. Appl. Conf.*, 2010, pp. 97–106.
- [47] D. Gupta and A. Etemad, "Privacy-preserving remote heart rate estimation from facial videos," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2023, pp. 706–712.

- [48] X. Jin, P. Y. Chen, C. Y. Hsu, C. M. Yu, and T. Chen, "CAFE: Catastrophic data leakage in vertical federated learning," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2021, pp. 994–1006.
- [49] M. Chen, X. Liao, and M. Wu, "PulseEdit: Editing physiological signals in facial videos for privacy protection," *IEEE Trans. Inf. Forensics Secur.*, vol. 17, pp. 457–471, 2022.
- [50] N. Zeller, F. Quint, and U. Stilla, "Calibration and accuracy analysis of a focused plenoptic camera," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 2, pp. 205–212, 2014.
- [51] W. Wang, A. C. d. Brinker, S. Stuijk, and G. d. Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, Jul. 2017.
- [52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [53] M. Schmidt, G. Fung, and R. Rosales, "Fast optimization methods for L1 regularization: A comparative study and two new approaches," in *Proc. Mach. Learn.: 18th Eur. Conf. Mach. Learn.*, 2007, pp. 286–297.
- [54] S. I. Serengil and A. Ozpinar, "LightFace: A hybrid deep face recognition framework," in *Proc. Innov. Intell. Syst. Appl. Conf.*, 2020, pp. 1–5.
- [55] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," in *Proc. Asian Conf. Comput. Vis.*, 2010, pp. 709–720.
- [56] S. W. Zamir et al., "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 14816–14826.
- [57] C. Downey, S. Ng, D. Jayne, and D. Wong, "Reliability of a wearable wireless patch for continuous remote monitoring of vital signs in patients recovering from major surgery: A clinical validation study from the tracing trial," *BMJ Open*, vol. 9, no. 8, 2019, Art. no. e031150.
- [58] L. Liu, D. Yu, H. Lu, C. Shan, and W. Wang, "Camera-based seismocardiogram for heart rate variability monitoring," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 5, pp. 2794–2805, 2024, doi: [10.1109/JBHI.2024.3370394](https://doi.org/10.1109/JBHI.2024.3370394).
- [59] G. Slapničar, W. Wang, and M. Luštrek, "Classification of hemodynamics scenarios from a public radar dataset using a deep learning approach," *Sensors*, vol. 21, no. 5, 2021, Art. no. 1836.
- [60] Z. Huang, W. Wang, and G. d. Haan, "Nose breathing or mouth breathing? A thermography-based new measurement for sleep monitoring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 3877–3883.
- [61] D. Huang, L. Ren, H. Lu, and W. Wang, "A scene adaption framework for infant cry detection in obstetrics," in *Proc. IEEE 45th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2023, pp. 1–5.