



MiSleep: Human Sleep Posture Identification from Deep Learning Augmented Millimeter-wave Wireless Systems

AAKRITI ADHIKARI and SANJIB SUR, University of South Carolina, Columbia, USA

In this work, we propose *MiSleep*, a deep learning augmented millimeter-wave (mmWave) wireless system to monitor human sleep posture by predicting the 3D location of the body joints of a person during sleep. Unlike existing vision- or wearable-based sleep monitoring systems, *MiSleep* is not privacy-invasive and does not require users to wear anything on their body. *MiSleep* leverages knowledge of human anatomical features and deep learning models to solve challenges in existing mmWave devices with low-resolution and aliased imaging and specularity in signals. *MiSleep* builds the model by learning the relationship between mmWave reflected signals and body postures from thousands of existing samples. Since a practical sleep also involves sudden toss-turns, which could introduce errors in posture prediction, *MiSleep* designs a state machine based on the reflected signals to classify the sleeping states into rest or toss-turn and predict the posture only during the rest states. We evaluate *MiSleep* with real data collected from Commercial-Off-The-Shelf mmWave devices for eight volunteers of diverse ages, genders, and heights performing different sleep postures. We observe that *MiSleep* identifies the toss-turn events start time and duration within 1.25 s and 1.7 s of the ground truth, respectively, and predicts the 3D location of body joints with a median error of 1.3 cm only and can perform even under the blankets, with accuracy on par with the existing vision-based system, unlocking the potential of mmWave systems for privacy-noninvasive at-home healthcare applications.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools;** • **Computing methodologies** → **Machine learning approaches;**

Additional Key Words and Phrases: 5G, millimeter-wave, sleep posture monitoring, HMM-Viterbi, convolutional neural networks

ACM Reference Format:

Aakriti Adhikari and Sanjib Sur. 2024. MiSleep: Human Sleep Posture Identification from Deep Learning Augmented Millimeter-wave Wireless Systems. *ACM Trans. Internet Things* 5, 2, Article 9 (March 2024), 33 pages. <https://doi.org/10.1145/3643866>

1 INTRODUCTION

Humans spend approximately one-third of their lives sleeping. High-quality sleep is of vital importance for the short-term proper functioning of the human body and for long-term good health [1]. Chronic sleep deprivation, such as regularly sleeping less than the recommended 7–8 hours, has been associated with multiple health disorders and risks in cardiovascular, respiratory, neurological, gastrointestinal, immunological, dermatological, endocrine, and reproductive systems [2, 3].

This work is partially supported by the NSF under grants CNS-1910853, CAREER-2144505, and MRI-2018966.

Authors' address: A. Adhikari and S. Sur, University of South Carolina, 550 Assembly St, Columbia, SC 29201, Columbia, USA; e-mails: aakritit@email.sc.edu, sur@cse.sc.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2577-6207/2024/03-ART9

<https://doi.org/10.1145/3643866>

As per the national surveys [4, 5], only 25 to 50% of adults in the United States slept the recommended 7–8 hours, and 20 to 35% reported consistent sleep difficulties. As per the international survey [4], up to 40% of adults show sleep difficulties. These surveys indicate that poor sleep health is prevalent at both local and global scales. Due to the well-recognized importance of sleep and the high prevalence of its inadequacy, sleep monitoring has become an important research area. *A key metric to monitoring sleep is the spatial and temporal understanding of sleep postures through the night*, as the postures directly influence sleep behavior and critical parameters [6–8]. Each of us sleeps in one of the broad categories of posture, such as supine, lateral, fetal, and so on, and exhibits wide variations of them throughout the night [9, 10]. The effect of different sleep postures has been studied widely to identify its relationship to different health conditions, such as brain activities [11], nocturnal bruxism [12], carpal tunnel syndrome [13], progressive glaucoma [14], motor development in infants [15], back pain in adults [16], waking cervical [17], and scapular and arm pain [18].

Specific sleep postures could also be fatal, depending on the pre-existing medical conditions. For example, supine posture is linked with exacerbating obstructive sleep apnea by creating unfavorable airway geometry, causing a reduction in lung volume and limiting the movement of airway dilator muscles, which could be life-threatening [19]. Infrequent turns due to impairment in control of the motor activity of Parkinson's patients lead to parasomnia and restless leg syndrome [20]. Prone sleep posture is associated with nearly 73% of sudden unexpected deaths of Epilepsy patients [21]. Infrequent changes in sleep posture are also the primary cause of pressure ulcers (i.e., bedsores) in post-surgical and elderly patients. Additionally, physicians recommend different sleep postures for different medical conditions: It is recommended to sleep on the side posture to reduce snoring, or left side to prevent heartburn, or supine posture to lower back or shoulder pain, or fetal posture during pregnancy, or some specific posture variations during post-surgery recovery [22–25]. *These examples highlight the importance of a sleep posture monitoring system that can provide real spatio-temporal observations, which could not only help with corrections but also prevent fatal accidents.*

Since it requires time to train a patient to adapt to a new sleep posture, physicians may need to frequently observe the fine-grained posture, such as skeletal information, and its changes throughout the night. Apart from error-prone qualitative assessment, where doctors ask patients (or their caretaker/partner) about their sleep postures, in-clinic quantitative assessment relies on visually observing the posture or inferring them by analyzing physiological signals from devices attached to a patient, such as Electrocardiography monitors, accelerometer, and so on [26–28]. Such assessments are typically done in a hospital setting, requiring patients to stay there overnight [29]. Recent works aimed to provide at-home monitoring [27, 30, 31], but they mostly require users wearing sensors or placing the sensors on the bed. Wearables are not only cumbersome during sleep but also unreliable, since certain patients, such as the elderly, tend to forget to wear them. Pressure sensors attached to the bed mattress are an alternative solution [7, 32–34], but they are costlier and often bring sleep discomfort. Contactless sleep monitoring, such as vision-based solutions, relies on optical and depth cameras to accurately monitor sleep postures, but they are highly privacy-invasive, and their performance is hindered by dark bedroom conditions and occlusion due to blanket [35–38]. Wireless-based solutions can overcome these challenges by inferring postures under no light without being privacy-invasive [39–41], but existing solutions rely on special-purpose low-frequency devices. Besides, they can only classify sleep postures into broad, discrete categories [42] and are unable to provide fine-grained posture information, such as the location of different body joints.

Fortunately, high-frequency **millimeter-wave (mmWave)** wireless devices provide an effective alternative to the existing systems to enable fine-grained posture monitoring: MmWave signals

can penetrate certain obstacles, work under zero visibility, and have higher-resolution than Wi-Fi. So, mmWave imaging can facilitate “seeing” the body posture under dark conditions and under the blanket. Besides, *mmWave transceivers are poised to soon become ubiquitous in all 5G-and-beyond smart home devices, such as routers and access points, enabling the opportunity for bringing privacy non-invasive human sleep posture monitoring system to the masses at home.* However, there exist two fundamental challenges in mmWave imaging. *First*, mmWave signals could be absorbed by many body parts or specularly reflect from them in different directions, away from the device, causing most signals to never reach back to the receiver [43]. So, the output human shape could have a lot of missing parts from which it is difficult to infer joint locations. *Second*, mmWave devices have extremely low resolution compared to vision-based systems; so, many high-frequency components, such as the contour and limbs, will be eliminated from the generated images [44]. Moreover, the reflected signals carry additional information about the bed and surrounding objects close to the body, making it harder to separate the human shape. So, it is challenging to extract body joint information and changes directly from traditional mmWave imaging during sleep.

To overcome these challenges, we propose MiSleep, a single-person sleep posture monitoring system that leverages signal processing and deep learning models to enable fine-grained monitoring continuously and non-intrusively with commodity mmWave devices. Instead of generating a mmWave image from traditional algorithms and then predicting the body joint locations, *MiSleep* directly predicts the joint locations from the reflected mmWave signals by learning the hidden association between them from thousands of data samples. To learn such an association, *MiSleep* employs a customized **Deep Convolutional Neural Network (DCNN)** that predicts the 3D locations of several key body joints from the reflected signals captured by multiple mmWave antennas. The reflected signals carry amplitude and phase information at various azimuth angles, elevation angles, and depth, and the DCNN extracts relevant features from multiple spatially co-located antennas’ signals to formulate a mapping of the signals to the body joints. Furthermore, to generalize the model for diverse populations, *MiSleep* models a height classifier and uses the error in its prediction to fine-tune the model. We use a dataset collected from several static sleep postures from multiple volunteers, and at runtime, *MiSleep* can predict 3D joint locations directly from the mmWave signals. However, the reflected mmWave signals could be corrupted by various factors, such as the Doppler effect, under the toss-turn during sleep. For example, a person could turn from a lateral to a supine posture in the middle of the night. Predicting the body posture, with a model trained on static postures, during such sudden movements not only is challenging but also is less useful, since toss-turns usually span for a short duration of a few seconds. Therefore, *MiSleep* designs a toss-turn detection module that can first classify the sleeping states to either rest or toss-turn. Then, it predicts the joint locations only during the rest state.

We design and prototype *MiSleep* with **Commercial-Off-The-Shelf (COTS)** devices by building a customized setup with two 77–81 GHz mmWave transceivers [45] to collect the reflected signals and a Microsoft Kinect Xbox One [46] to collect the ground truth 3D joint locations. We collect reflected signals and ground truths of nearly 70k samples (>18.5 GB) from multiple experiments under different conditions with nine volunteers. The experimental results show that *MiSleep* can detect all ground truth toss-turn events and can identify the start time and duration within 1.25 s and 1.7 s of the ground truth, respectively, for all cases. For static sleep postures with a base model, *MiSleep* predicts the 3D location of body joints with a median error of 1.3 cm only. Furthermore, *MiSleep* generalizes well for diverse volunteers with median and 90th percentile errors of 2.3 cm and 7.4 cm, respectively. Even when we test performance on unseen volunteers, we observe *MiSleep* provides acceptable joint prediction. Furthermore, *MiSleep* can identify the body postures and locate the joints under blankets with similar accuracies. Finally, *MiSleep* can classify volunteers by leveraging their height information with an accuracy of 87% to 94% for eight

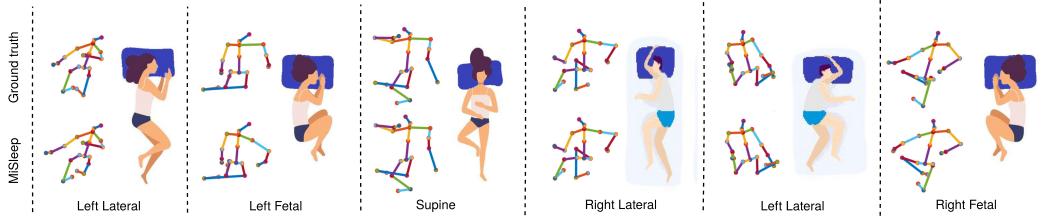


Fig. 1. *MiSleep* predicts the location of key body joints during sleep using deep learning on millimeter-wave wireless signals from off-the-shelf devices.

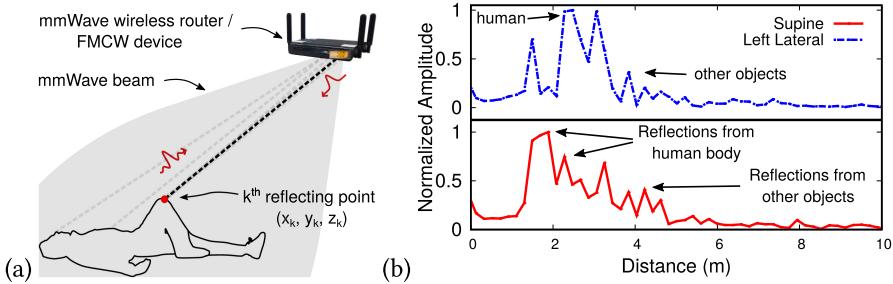


Fig. 2. (a) A mmWave device captures reflected signals from the target. (b) Example of reflected signals from two sleep postures.

volunteers and can classify five broad categories of postures with an accuracy above 90% for all volunteers.

In summary, we make these two contributions: (1) We design a customized deep learning framework for predicting the 3D location of body joints during sleep from COTS mmWave devices, which generalizes well for a diverse population. To the best of our knowledge, *MiSleep* is the first system to infer sleep postures in the form of 3D joint locations from the COTS mmWave device and achieve accuracy on par with the existing vision-based systems. (2) We design a toss-turn detection module that can accurately identify key sleep events and their timing information from the mmWave reflected signals. To accelerate the research on COTS mmWave device-based sleep monitoring, we will open-source our dataset and codebase.

2 BACKGROUND AND CHALLENGES

2.1 Millimeter-wave Reflections and Imaging

Traditional mmWave imaging approaches rely on **Frequency Modulated Continuous Waves (FMCW)** from a device to generate an image [47]. The device illuminates the target scene with a wideband and wide-beamwidth FMCW signal (Figure 2(a)). Each FMCW signal sweeps one of the mmWave frequency bands linearly (e.g., 77 to 79 GHz, where 2 GHz is the signal bandwidth) and receives the signals reflected back from various objects in the surroundings, including the sleeping human (see Figure 2). Let us assume the human body is composed of n reflecting points at different distances, d_1, d_2, \dots, d_n , from the transceiver. Since points at different distances will reflect back signals at slightly different times, the receiver can identify the amplitude and phase of the signals by measuring the difference of frequencies between the received and transmitted signal to obtain the signal **time of flight (ToF)**. The ToF is then translated into object distance based on $d_i = c \cdot \text{To}F_i/2$ [47], where c is the wireless propagation speed. The signals received at time t

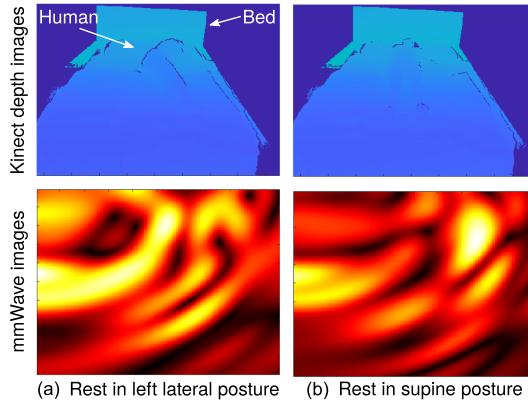


Fig. 3. (Top) Kinect depth images for two sleep postures: Left lateral; and Supine. (Bottom) mmWave images are generated from a device with three transmit and four receive antennas.

can then be expressed as a complex vector of length n , $R_t(\{d_1, d_2, \dots, d_n\})$. Here, $d_{i+1} - d_i = c/2B$, where B is the signal bandwidth. Figure 2(b) shows two examples of such reflected signals for two sleep postures.

While a practical networking device does not possess the capability of the FMCW transceiver, such reflected signals can still be obtained from them by estimating the **Channel Impulse Response (CIR)** from communication packets. For example, all IEEE 802.11ad networking devices, operating at the 60 GHz mmWave frequency band, allow measuring such CIR from the **Channel Estimation (CE)** header field of each standard packet [44, 48]. But the signal received by a single antenna cannot be used to generate an image, since the target scene is in 3D and reflected signal is in 1D. So, the transceiver uses either the reflections from multiple antennas or sweeps a narrow beam along the horizontal and vertical directions to locate different reflecting points at relative azimuth and elevation angles. Then, based on the angular and distance information, the reflection points can be mapped into the 2D plane to generate a mmWave image. The image resolution in depth, azimuth, and elevation depends on the signal bandwidth, number of horizontal antennas, and number of vertical antennas, respectively [49, 50].

2.2 Challenges with Direct mmWave Imaging-based Posture Detection

Identifying sleep postures directly from the mmWave images, however, is challenging for multiple reasons. *First*, existing commercial devices are usually designed with a small number of antennas, such as four or eight, in the horizontal and vertical directions, [51–55]: So, the generated mmWave images will have an extremely low resolution. While such images could be potentially fed into a deep learning model to classify broad categories of postures, it is hard to identify any body joint locations from them. *Second*, future mmWave devices could possess a large number of antennas, such as 1,024, in both horizontal and vertical directions [56, 57], which could improve the fundamental image resolution. However, all the antennas must be placed by strictly following the Nyquist spatial criterion to generate alias-free images. (The Nyquist criterion states that the distance between adjacent antenna elements should be $\sim c/(2f)$, where c is the propagation speed and f is the carrier frequency [58].) Adding the reflected signals from non-uniformly spaced antennas will create spurious reflection points, and the image will appear distorted [59]. Figure 3 shows examples of such distortions, where the mmWave images have no visual correlation with the actual scene. While future devices could include a large number of antennas, the antennas will be placed non-uniformly across the device for a 360° network coverage. So, they will produce aliased images

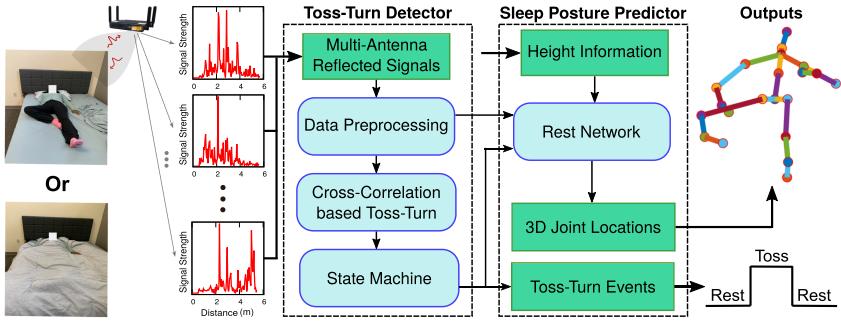


Fig. 4. System overview of *MiSleep*.

with unrecognizable body joint locations. Finally, different body parts reflect mmWave signals differently, such as the torso could reflect a strong signal, but the limbs usually reflect weak signals [44, 60]. Also, the absolute strength of the reflections from the same body part varies over time if the posture has changed. This is because some of the body parts could reflect the signal specularly, so depending on its relative orientation w.r.t. the antenna, it might deflect the signal away from the device [61]. Due to such uneven reflections, at a specific time instant, only a subset of body parts is visible to the mmWave device, making it challenging to identify the location of key body joints.

3 MISLEEP DESIGN

3.1 Overview

MiSleep aims to bring a continuous, non-intrusive, and non-privacy-invasive sleep monitoring system at home by leveraging COTS mmWave devices. Recording an accurate 3D locations of body joints throughout the night could enable numerous applications, such as baseline monitoring of patients, sleep diary to assist physicians, classification of sleep posture, track changes of body parts, toss and turn detection, detection of sudden movement during the night, amount of time a person is asleep or is awake or is restless, and so on.

To this end, *MiSleep* uses mmWave transceivers to illuminate the target scene with an FMCW signal and subsequently receives the signals reflected back from various objects in the surrounding environment, including the sleeping human. After capturing these reflected signals, *MiSleep* applies **Fast Fourier Transform (FFT)** to them, where the location of peaks in the frequency spectrum corresponds to the range of objects. The received signals, after undergoing FFT, serve as the input to our system, referred to as mmWave reflected signals. Then, *MiSleep* designs two modules: A toss-turn detector that classifies the sleeping period into two states, rest or toss-turn, and a sleep posture predictor that predicts the 3D location of body joints during the rest state. For the toss-turn detection, *MiSleep* leverages the cross-correlation between successive mmWave reflected signals and a **Hidden Markov Model (HMM)** to label the sleeping period. For the sleep posture prediction, instead of relying on traditional imaging algorithms, *MiSleep* trains a customized deep learning framework with thousands of examples of mmWave signal reflections and ground truth 3D location of body joints to learn a generalized relationship between them. Then, during the runtime, *MiSleep* can accurately predict the joint locations only from the mmWave reflected signals. We collect diverse datasets with multiple volunteers performing different sleep postures, which record signal reflections from the body and the ground truth 3D location of the joints from co-located mmWave transceivers and an RGB-D camera, respectively. These datasets are then fed to *MiSleep* for training a height agnostic **Deep Convolutional Neural Network (DCNN)**. The

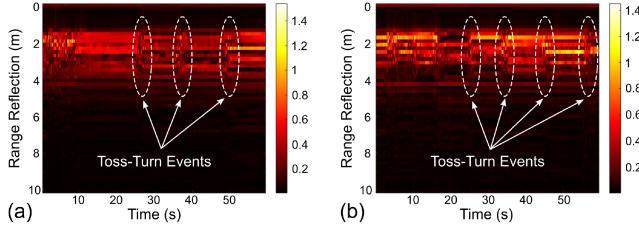


Fig. 5. (a–b) STFT outputs from two monitored cases with three and four toss-turn events. Reflected signals are collected at 40 ms intervals for 60 seconds, and events are marked by visually inspecting the depth images.

mmWave reflected signals are paired with the ground truth joint locations and the known height of the monitored user to train the network. The DCNN, from thousands of data pairs, learns the association between reflections and joint locations, and at runtime, predicts the joint locations only from the reflected signals. Figure 4 shows the system overview of *MiSleep*. We now describe these design components in detail.

3.2 Data Synchronization and Resampling

MiSleep's deep learning model relies on datasets collected from different COTS devices that do not have a tight hardware synchronization or the same sampling rates. Therefore, it is critical to ensure synchronization between them so input-output data pairs are aligned for training. To this end, we rely on software synchronization and process data to remove any residual misalignment. We collect the UTC timestamp from an NTP timeserver before triggering the mmWave devices and the RGB-D camera for data collection. Then, based on timing information, we correlate the first received frame with all other frames in mmWave devices, which identifies the first local timestamp of movement and allows synchronization between devices. Once we find the first local timestamp of movement, we calibrate data samples by offsetting the samples w.r.t. the timestamps. Additionally, to compensate for the sampling rate mismatch (the mmWave devices and RGB-D camera in our setup have 25 and 30 fps sampling rates, respectively), we resample the data in time using a weighted averaging of adjacent samples. Then, the processed datasets are fed into the toss-turn detector to classify the sleep states as rest or toss-turn, and under the rest states, they are fed into the sleep posture detector to detect the 3D location of body joints.

3.3 Toss-turn Detection and State Machine

The core purpose of the toss-turn detector module is to identify sudden movements during sleep and classify the sleeping period into two states: Rest or toss-turn. It is critical to identify and separate the states, as it not only helps in estimating the time gap between two resting periods but also facilitates a deep learning model to only predict the posture under rest states and avoids erroneous predictions under toss-turns. Furthermore, it helps *MiSleep* understand that the person has likely changed the sleep posture so the posture predictor module can reset itself. Besides, identifying and recording such states could be beneficial for long-term monitoring. To classify the states, *MiSleep* analyzes and processes the reflected mmWave signals received by the monitoring device.

3.3.1 Cross-correlation-based Toss-turn Detection. Inspired by the previous works on elderly fall detection using wireless signals [62, 63], *MiSleep* leverages the observation that *in comparison to the rest states, toss-turn states are usually associated with significantly higher spatio-temporal changes in the received mmWave signals*. While it might appear logical to detect toss-turn states based on the velocity changes of the reflection points, as movement often translates to a change

in velocity, during sleep, individuals often exhibit minor, sometimes sudden, movements without necessarily shifting their overall posture. For instance, someone might momentarily twitch an arm or leg or adjust their head position. Such sporadic movements, though seemingly minor, can result in noticeable velocity changes in the reflecting points. By solely relying on these velocity changes, our system could misconstrue these transient movements as significant toss-turn events.

To this end, we perform a preliminary experiment with our setup with a bed placed at 2.5 m in front of the mmWave devices (see Figure 12), and a volunteer is asked to lie down and perform multiple toss-turns, i.e., move from one posture to another. We collect the reflected mmWave signals as a frame at 40 ms intervals from the experimental setup for 60 seconds and apply a **Short-Time Fourier Transform (STFT)** over the signal received by one of the mmWave antennas. Figures 5(a-b) show the STFT output for two cases, with three and four toss-turn events. While some changes are observed under the toss-turn states, it is hard to separate them from the rest states without consulting the ground truth. The changes are much weaker than the stark time-and-frequency changes observed in the wireless signals during the fall [62]. A fall is a large-scale event associated with a change in the height, the centroid of the body, and distance of the limbs and torso from the monitoring device, which creates a strong, unique signature in the reflected signals. In comparison, the toss-turn during sleep is usually a small-scale event, where the centroid of the body might not change, and the limbs typically move between adjacent distance bins of the reflected signals (Section 2). Therefore, it will require additional signal processing to amplify the changes during the toss-turns and separate them from the rest states.

To amplify such changes, *MiSleep* applies a cross-correlation between successive frames of the reflected signals and estimates the rate of change (i.e., time-derivative) in the peak correlation output. The key idea is intuitive: Cross-correlation between successive frames allows uncovering the similarity (or dissimilarity) between the consecutive reflected signals. Since, during the rest states, there are almost no changes in the successive reflected signals, the cross-correlation will show almost the same peak; so, its rate of change over time should be close to zero. However, during the toss-turn states, such correlation peak fluctuates significantly, with a variable rate of change. More importantly, the time-derivative removes the almost constant reflections from the static background, i.e., bed, furniture, nightstands, and so on, so the only changes due to the body movement are amplified and stand out. Let us consider $R_t(\{d_1, d_2, \dots, d_n\})$ as the reflected signal at time t from distances d_i w.r.t. the receiver. Mathematically, the cross-correlation, $xCorr$, and its time-derivative, $\Delta xCorr$, can be expressed as follows:

$$xCorr_t = \max |R_t(\cdot) * R_{t-1}(\cdot)| = \max_{m \in (0, n-1)} \left| \sum_{i=1}^{n-m} R_t(d_{i+m}) \cdot R_{t-1}^*(d_i) \right|$$

$$\Delta xCorr_{t-1}^t = xCorr_t - xCorr_{t-1},$$

where R_{t-1}^* is the complex conjugate of the received reflected signal at time instant $t-1$. Figure 6 shows an example of monitoring for ~ 30 seconds with four toss-turn events, the corresponding cross-correlation output, and its time-derivative, which shows a stark contrast between the toss-turn and rest states. To reduce the number of oscillations between the false detections (+/-) and states, *MiSleep* smoothens $\Delta xCorr$ over time with an envelope detector using a Hilbert Transformation, similar to Reference [64]. The envelope detector uses the **Root-Mean-Square (RMS)** of $\Delta xCorr$ amplitudes over N consecutive frames. Intuitively, a large value of N suppresses many false detections but will have a slow reaction to the true state change. However, a small value of N will have a fast reaction but could lead to high false detections and state oscillations. In practice, $N = 25$ frames, i.e., 1 second of the consecutive reflected signals, for envelope estimation, yields a good result, since human movements during the sleep are on the order of several seconds [65].

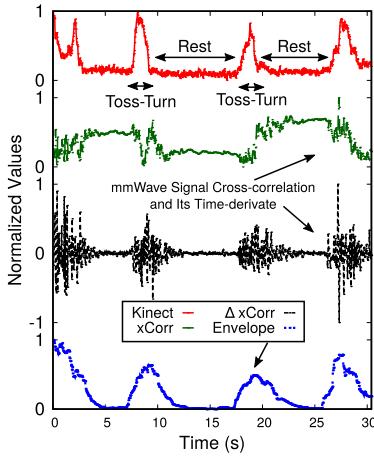


Fig. 6. Example toss-turn/rest states detected from Kinect, cross-correlation between reflected mmWave signals, its time-derivative, and envelope estimation with RMS of samples for 1 second.

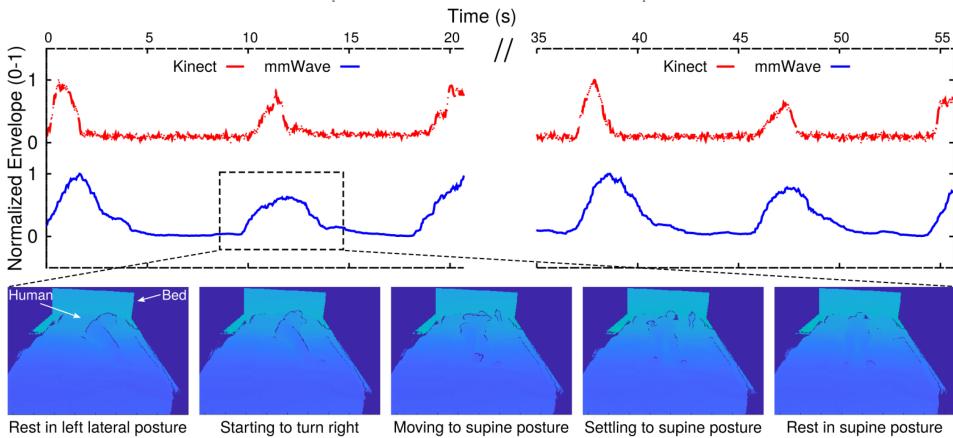


Fig. 7. Monitoring toss-turns: Envelope detector output from mmWave signals follows the changes in posture.

Figure 7 shows another output of the envelope detector and compares the result with the Kinect-based output. It also shows a zoomed time period, where a volunteer in the left lateral posture turns right and moves to the supine posture.

► **Practical Challenges:** There are two practical challenges in using the envelope output for state detection. *First*, since *MiSleep* uses a wideband mmWave device, the signals may occasionally show strong fluctuations in the correlation output, even under the rest state, which can lead to false detections. This is because the signals include reflections from all objects in the scene, even far away. For example, with 256 samples in each frame of the reflected signals and 1.62 GHz bandwidth, *MiSleep* is capable of measuring reflections up to 23.69 m. Clearly, dynamic events (such as another human walking in another room) from such a faraway distance not only are irrelevant for short-range monitoring but also can affect state detection. So, *MiSleep* considers a maximum range and prunes the reflected signals that are beyond the threshold. In our design and evaluation, we use

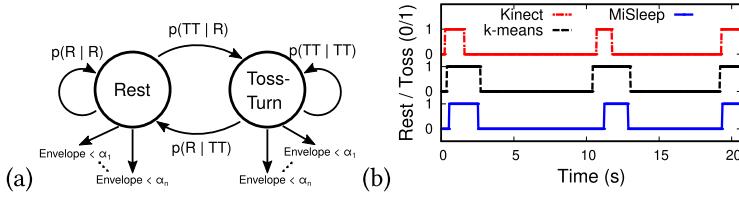


Fig. 8. (a) A two-state HMM. We use the Kinect output to learn the state and emission matrices and then use a Viterbi decoder on the envelope to predict the states. (b) Example detection of rest and toss-turn states.

a maximum range of 5 m to cover only the bed and the surrounding area. *Second*, the output of the envelope detector is a real number between 0 and 1, where 0 indicates no change in the successive reflected signals, and 1 indicates a very high change. But for posture detection, *MiSleep* requires discrete binary states: rest or toss-turn. A strawman approach to discretize the envelope could be to use a fixed threshold, such as $envelope < 0.5$ as rest, and $envelope \geq 0.5$ as toss-turn; or, apply an adaptive threshold with a temporal clustering algorithm, such as *k-means*, over small observation periods. But such approaches can often lead to false detections. This is because mmWave signals are very sensitive to the minute changes in the environment; so, the envelope detector can still show high output during the occasional hand or leg movements, even if the full body has not turned yet. Also, there could be early toss start and late toss end detection, leading to the wrong estimation of the event duration: See Figure 7, where the turn-time duration estimated by mmWave signals is much larger than the Kinect output.

3.3.2 Improving Detection with a Two-state HMM. To overcome this challenge and improve the detection accuracy and timing estimations, we design a lightweight two-state HMM [66]. The HMM not only converts the envelope with real-valued output between 0 and 1 to a discrete output of 0 and 1, but also improves the state detection accuracy and reduces the timing errors. Figure 8(a) shows the state transition diagram of *MiSleep*'s HMM: The two states are rest and toss-turn, and the emissions are different levels of envelope values.

To build the HMM, we collect several ground truth datasets involving multiple volunteers tossing and turning during their sleep and formulate the state transition matrices by estimating the four conditional probabilities, i.e., $p(Rest|Rest)$, $p(Rest|Toss-Turn)$, $p(Toss-Turn|Rest)$, and $p(Toss-Turn|Toss-Turn)$. We formulate the emission matrix by estimating the conditional probabilities for discrete envelope values (e), i.e., $p(e < \alpha_1|Rest)$, $p(e < \alpha_2|Rest)$, ..., $p(e < \alpha_1|Toss-Turn)$, $p(e < \alpha_2|Toss-Turn)$, and so on. Finally, at runtime, *MiSleep* first calculates the envelope from the reflected mmWave signals and then uses the state transition and emission matrices and a Viterbi decoder [66] to predict the binary states, corresponding to rest and toss-turn. Figure 8(b) shows an example of ~20 seconds of monitoring with three toss-turn events and compares the prediction with the ground truth Kinect-based output. Clearly, in comparison to the *k-means* with adaptive threshold, HMM can improve the errors in event start and stop times. Once the entire sleeping period is classified into rest and toss-turn states, *MiSleep* aims to predict the sleep posture during the rest state.

3.4 Deep Learning-based Sleep Posture Prediction

MiSleep predicts sleep posture using a deep learning model that relies on the relationship between the 3D location of body joints and the mmWave reflected signals. The model can only learn such a relationship based on the feature variance between distinct postures and individuals. So, we first analyze the behavior of reflected signals and their relationship with individuals' sleep postures.

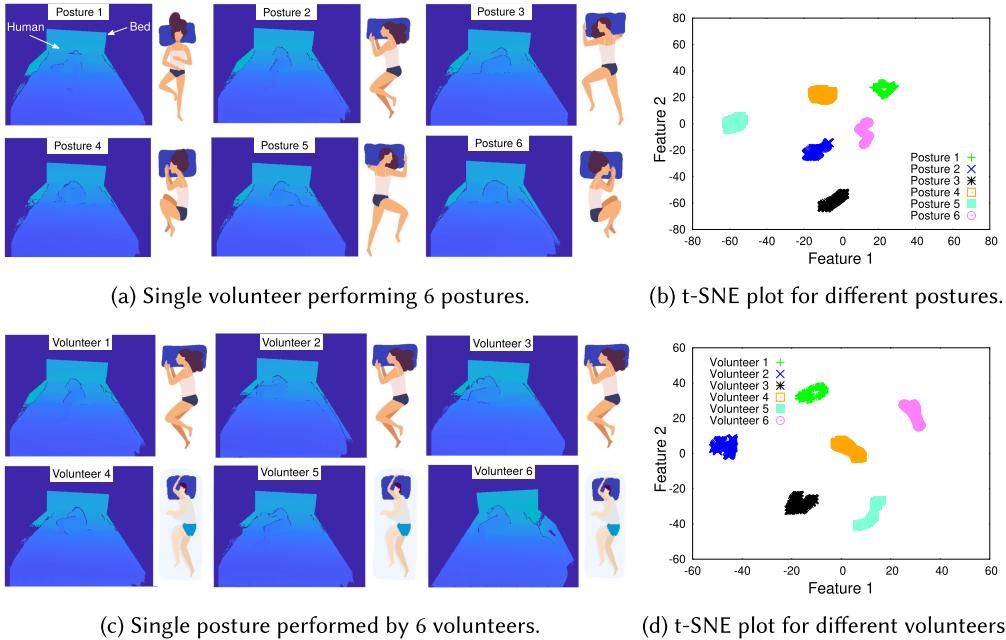


Fig. 9. t-SNE analyses of reflected signals from different sleep postures and volunteers show unique clusters in the 2D feature space.

3.4.1 Relationship between Human Sleep Postures and Signal Reflections. Intuitively, we can predict the 3D location of joints for a specific posture only if the raw reflected signals from various postures from the same human demonstrate distinct behavior in feature space. Similarly, we can distinguish the 3D location of joints of individuals performing the same posture only if raw reflected signals from a different person (i.e., varying in height and gender) appear distinct in feature space. These distinctions not only will help the model to capture the relationship between raw reflected signals and 3D joint locations but also are useful in generalizing the model for a diverse population. To this end, we conduct two sets of experiments: *First*, we collect mmWave reflected signals from our setup with the bed from a single volunteer performing six different sleep postures (Figure 9(a)). *Second*, we ask six volunteers (three males and three females, height varying from 155 to 178 cm) to lie down on the bed in the same posture (Posture 2 in Figure 11) and collect signals reflected from their bodies (Figure 9(c)). For each experiment, we project the reflected signals in two-dimensional feature space by measuring the t-SNE distribution; this distribution represents the signals in such a way that the input with similar features appears closer to each other.

Figures 9(b) and 9(d) show the t-SNE distribution for a single volunteer performing six different sleep postures on the bed and for six volunteers performing the same posture, respectively. Clearly, we observe six unique feature clusters for both the cases: Such unique clusters indicate that the input signal carries enough signature so a learning model should be able to effectively learn and distinguish them using mmWave reflected signals.

3.4.2 MiSleep’s Rest Network. The core purpose of the Rest Network is to predict the 3D location of body joints from the mmWave signal reflections and capture diverse sleep postures during the rest state. The Rest Network is designed using a customized DCNN called Joint Regressor to map the relevant higher-dimensional features in input to output. The Joint Regressor is trained with two human-anatomy specific features. *First*, 3D location of body joints of an individual is correlated

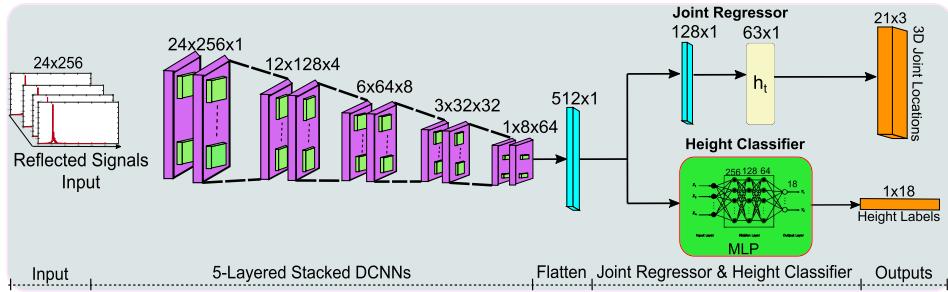


Fig. 10. *MiSleep*'s Rest Network architecture.

with her height [67, 68]; so, *MiSleep* could constrain and fine-tune the prediction for joint locations by predicting the height and comparing the difference with the known height of the user. Then, the model can output better 3D joint locations by backpropagating the height prediction error, and the network can be generalizable for many users. *Second*, most of the human body joints are spatially connected to each other in a *parent-child*, tree-like hierarchy [69], and 3D pose of one joint is usually constrained by its parent's pose [70]. So, the 3D location output of a child joint should be conditioned on its parent joint to ensure the distance between the parent and child is always fixed, across all postures. We now briefly discuss the DCNN fundamentals and then describe the network components in detail.

► **DCNN as the Feature Extractor:** A DCNN maps relevant features in input to output by using filters through a series of the convolution operation: It extracts the spatial features relevant to the network by observing the non-linear correlation between input-output pairs. A DCNN consists of input and output layers and a series of hidden layers along with millions of parameters that allow it to model non-linear relationship, which is otherwise difficult to observe through mathematical regression models [71]. However, DCNN needs to learn the mapping between joints and reflections as well as learn interconnected spatial dependencies between joints. Therefore, DCNN needs deeper customization in its architecture. Inspired by the approach in DeepPose [72] where they propose a cascade of DCNN regressors, we propose to instead exploit the parent-child relationship between joints. If we exploit the parent-child relationship between joints, then we can model the joint hierarchy that governs the joint locations and can reduce the overhead on DCNN. This allows DCNN to focus on mapping signals to individual joints and propagate information into the output layer.

► **Rest Network Architecture:** Figure 10 shows the Rest Network architecture, with two major components: Joint Regressor and Height Classifier, which we discuss next.

Joint Regressor: The objective of the Joint Regressor is to capture the hidden relationship between mmWave reflections and 3D joint locations to infer the complete posture by using a customized DCNN as the Feature Extractor. A DCNN maps relevant features in input to output by using filters through a series of the convolution operation: It extracts the spatial features relevant to the network by observing the non-linear correlation between input-output pairs [71]. Joint Regressor's DCNN takes a 2D input and performs a series of 2D convolutions in several layers to learn the relationship between input and output. Joint Regressor is composed of several layers to first learn the basic features, and as it gets deeper, it learns deeper hidden features that map non-linear relationship between input and output. For the purpose of mapping signals to joint locations, we observe through a series of fine-tuning processes that five layers of stacked convolution with two convolution layers in each stack yield the best result than a vanilla DCNN. Stacked representation provides depth to the network so it can learn complex hidden representations [73]. We also apply

Table 1. Joint Regressor Network Parameters

Stack	C1 S1,S2	C2 S1,S2	C3 S1,S2	C4 S1,S2	C5 S1,S2	FL	FC1	FC2	Output
Filter #	4	8	16	32	64	512	128	63	21x3
Filter Size	6x6	6x6	6x6	6x6	6x6				
Dilation	2x2, 1x1								
Act. Fcn.	LReLU	LReLU	LReLU	LReLU	LReLU	LReLU	LReLU	Linear	

C_j: jth Convolutional layer; S_i: ith Stack in jth Convolutional layer; FL: Flatten layer; FC: Fully Connected layer; Act. Fcn.: Activation Function. LReLU: Leaky ReLU.

batch normalization after each stacked layer to ensure normalization and prevent overfitting. The five 2D stacked convolutional layers are connected to a flatten layer that converts the input to a 1D abstract feature of size 512 and then pass it through two fully connected layers of size 128 and 63, respectively, to finally give output as the 3D location of 21 joints. Table 1 shows the detailed network parameters.

Height Classifier: The objective of the Height Classifier is to assist the Joint Regressor in learning the association between diverse postures of the same person. Since the skeleton of a person typically depends upon her height [74, 75], incorporating height information can make the model generalizable to many individuals with very little or no fine-tuning. A user could input her ground truth height to the monitoring system, and *MiSleep* can constrain the output from the Joint Regressor by comparing the predicted height w.r.t. the ground truth and backpropagating the error to rectify the prediction of joint locations. Instead of predicting the actual height, we employ a classifier by quantizing human heights into discrete values and then predicting the class labels associated with the quantization. The reason behind designing the model in such a way is twofold. *First*, it is relatively easier to achieve higher accuracy in predicting a class label than regressing exact height when we work with small samples from a diverse population. *Second*, since human heights are limited to a certain deterministic range (e.g., in the US, the average height ranges 163 to 179 cm [76]), it is well-suited to discretize them into range bins, instead of regressing a real value of height, where the network could suffer from out-of-range issues. To learn the association between height and sleep postures to mmWave signals, Height Classifier takes input from the flatten layer in the Joint Regressor and uses a **Multilayer Perceptron (MLP)** to output a height classification. MLP is a neural network with one or more hidden layers of neurons that are fully connected in each layer to learn the mapping between input and output. Similar to DCNN, hidden layers provide levels of abstraction to learn the mapping between input and output; but in contrast to DCNN, MLPs are fully connected in each layer, providing all combinations of the features of the previous layers rather than relying on local spatial coherence with a small receptive field. Additionally, the input to the Height Classifier is a vector, and MLP works well with such a vector representation. MLP in the Height Classifier comprising three hidden layers with 256, 128, and 64 neurons and an output layer with the number of neurons equal to the number of height classes. We apply ReLU activation in each layer and a Softmax activation in the output layer, which outputs probabilities associated with the labels, and we select the label with the highest predicted probability.

► **Total Loss Function:** We train the Joint Regressor and the Height Classifier jointly by designing a custom loss function, which is crucial to ensure that the network converges to an optimal value by updating and backpropagating weights [77]. For N number of total joints, the loss for the Joint Regressor is a combination of the Euclidean distance loss, $L_{ED} = \sqrt{\sum_{i=1}^N (J_{real}^i - J_{pred}^i)^2}$, between the predicted (J_{pred}^i) and ground truth (J_{real}^i) for i th joint locations and the parent-child

Table 2. Posture Classifier Network Parameters

Stack	C1 S1,S2	C2 S1,S2	C3 S1,S2	C4 S1,S2	C5 S1,S2	FL	FC1	FC2	FC3	FC4	Output
Filter #	4	8	16	32	64	512	128	64	32	16	5
Filter Size	6x6	6x6	6x6	6x6	6x6						
Dilation	2x2, 1x1		LReLU	ReLU	ReLU	ReLU	Softmax				
Act. Fcn.	LReLU	LReLU	LReLU	LReLU	LReLU						

C_j: jth Convolutional layer; S_i: ith Stack in jth Convolutional layer; FL: Flatten layer; FC: Fully Connected layer; Act. Fcn.: Activation Function. LReLU: Leaky ReLU.

distance loss, $L_{JH} = \sum_{i=1}^N |PCD_{real}^i - PCD_{pred}^i|$, that captures the joint hierarchy between predicted (PCD_{pred}^i) and ground truth (PCD_{real}^i) distance of i th joint. The loss function in the Height Classifier is a categorical cross-entropy loss, $L_{HC} = -\sum_{i=1}^K (y_i^{real}.log y_i^{pred})$, between the i th predicted (y_i^{pred}) and ground truth (y_i^{real}) label of height for K number of class labels, which provides a good quantitative measure in distinguishing probability distributions of discrete categories. The total loss can be expressed as $L_{Total} = \lambda_1 \cdot L_{ED} + \lambda_2 \cdot L_{JH} + \lambda_3 \cdot L_{HC}$, where λ_1 , λ_2 , and λ_3 are the hyperparameters that govern the contribution of each loss, and we will discuss their choice in Section 4.

3.4.3 MiSleep’s Posture Classification Network. The core objective of this network is to elevate the capability of MiSleep to classify different sleep postures. To this end, we design a Posture Classifier that can classify the sleep postures directly from the reflected mmWave signals. We modify the network architecture of MiSleep in Figure 10 by adding a classifier to the Rest Network to learn to map reflections to different sleep postures. Similar to the Height Classifier, the Posture Classifier takes the reflected signals of size 24×256 as input and then leverages five layered 2D Stacked DCNNs and four fully connected dense layers to predict classes. We apply batch-normalization and LeakyReLU activation function in all the convolutional layers, and ReLU activation function in all dense layers. We also apply a softmax activation at the output layer of size 5 for the categorical classification. The softmax activation assigns a probability (0 to 1) to each class, indicating the likelihood of input belonging to each class. Table 2 shows the detailed network parameters.

To train the network, we select the categorical cross-entropy as the loss function between actual probabilities and predicted probabilities of different categories [78]. We discuss the network training parameters in Section 4 and present evaluation in Section 5. Our classifier is primarily focused on our baseline postures, but it can be expanded to include more postures without requiring substantial changes in the layers or training with larger samples.

To summarize the design components, MiSleep first classifies the sleeping states as rest or toss-turn and then predicts the 3D location of the key body joints during the rest state from the reflected mmWave signals, which, in turn, helps to classify the sleeping postures.

4 IMPLEMENTATION

4.1 Hardware Platform

We implement and evaluate MiSleep by collecting real datasets from a customized hardware setup we built. Our setup includes two mmWave transceivers operating at the 77–81 GHz unlicensed mmWave frequency bands, TI IWR1443BOOST [45], that collect the mmWave reflected signals in real-time at a frame rate of 25 fps, and one RGB-D camera, Microsoft Kinect Xbox One [46], that collects the ground truth depth images and 3D location of body joints at a frame rate of 30 fps. A data capture module, TI DCA1000EVM [79], is attached to each device to capture the mmWave signals in real-time at a frame rate of 25 fps. The RGB-D camera captures depth images and outputs

Fig. 11. Five broad categories of sleep posture in *MiSleep*.

Table 3. Volunteer Information

Characteristics	Value
Total number	8
Age range (years)	19–30
Male/Female (%)	62.5/37.5
Height range (cm)	155–178

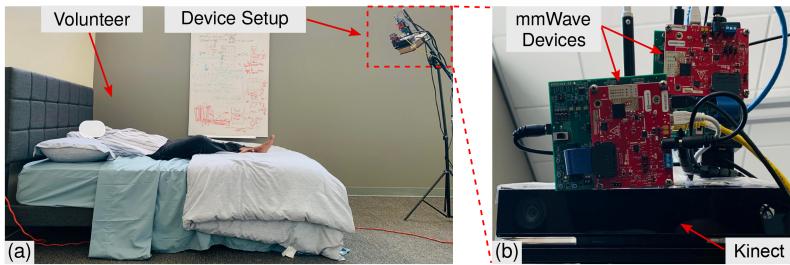


Fig. 12. (a) Experimental setup with the devices and the bed where a volunteer is performing a sleep posture. (b) Hardware platform with two 77–81 GHz mmWave devices and an RGB-D camera.

3D joint locations at a frame rate of 30 fps. We follow Section 3.2 to preprocess the datasets and synchronize and resample the captured signals and frames to align the datasets from multiple devices and further sanitize them to remove delays and spurious noise. Figures 12(a–b) show our experimental setup similar to at-home bedroom setting with a queen-size bed placed at 2.5 m from the monitoring equipment. Each mmWave device has three transmit and four receive antennas on a linear axis, and in its horizontal orientation, it can resolve reflection points in azimuth and depth, mostly. So, to resolve points also from elevation, we use two devices with one rotated 90° counter-clockwise w.r.t. another. The resultant setup has two antenna arrays arranged in 3×4 and 4×3 configurations, and with a bandwidth of 1.62 GHz, it achieves a depth resolution of 9.25 cm. To process the received signals, we apply traditional FMCW signal processing with the following parameters: Ramp start frequency – 77 GHz; frequency slope – 29.982 MHz/μS; baseband sampling rate – 1 Msps; number of ADC samples – 256; chirp sweep duration – 60 μS; pulse repetition rate – 1 kHz; and maximum antenna gain – 30 dB. Our input dimensions are 24×256, given by the number of ADC samples and two antenna arrays. Here, 24 corresponds to the count of virtual channels derived from both 3×4 and 4×3 configurations, while 256 denotes the number of ADC samples. We implement *MiSleep* on Matlab and Python environments running on host PC and GPU servers, which use the mmWave reflected signals as input and generate rest/toss-turn states, 3D location of body joints, and sleep posture class as outputs.

4.2 Real Data Collection

We collect datasets from eight volunteers (age: 19–30 years, M/F percentage: 62.5/37.5, height range: 155–178 cm) performing several variations of the five sleep postures on the bed: See Figure 11 and Table 3 for the sleep postures and volunteers' information. The background in our setup consists of drywalls, like a bedroom, without any clutter (except for the whiteboard on the right side).

For the toss-turn datasets, we ask a volunteer to lie down on the bed and perform multiple toss-turns, i.e., move from one posture to another, within 60 seconds. For static posture datasets, a single experiment requires 60 seconds to complete, where the first 7 seconds are spent on walking

into the setup and looking into the Kinect for about 5 seconds so Kinect starts detecting the joints. Then, the volunteer is required to lie down in a posture till the end of the experiment. Even though we collect data for 60 seconds with 30 fps, we discard many frames during preprocessing and only consider those frames where the body joints are successfully detected by the Kinect during the time the volunteer is in a static posture.

For height classification, we quantize the height range into 18 equal-sized bins from 147 cm to 191 cm and label them 1 to 18. We train the Height Classifier with these ground truth labels and fine-tune the Joint Regressor using the loss from this classifier (Section 3.4.2). For heights beyond this range, little fine-tuning can help the network to better understand the association between the reflected signals and body joints. In total, we have nearly 70k samples with a raw data size of 16.2 GB of toss-turn events, with a processed data size of 2.5 GB with five postures from eight volunteers of diverse ages, genders, and heights. *Such diverse datasets help MiSleep to not only build accurate models that are generalizable but also help benchmark and test them for multiple volunteers.*

4.3 Network Training

We train *MiSleep*'s Rest Network by exploring different parameter settings to ensure convergence to a near-optimal value. *MiSleep*'s Rest Network predicts two outputs, 3D joint locations and height class, and trains the network end-to-end with different loss functions for each output. *First*, we set the training epochs as 2,500 and then monitor the training process by saving weights of the best model using model checkpoint and callbacks till the total loss function shows no improvement for consecutive 30 epochs. *Then*, we explore different optimizers, such as, *Adam*, *Rmsprop*, *SGD*, and so on, and observe better convergence with *Adam* with a learning rate of 2×10^{-4} and a batch size of 2. To ensure better convergence and prevent overfitting, we split the training set into training and validation sets in an 8:2 ratio. *MiSleep*'s Rest Network includes three different losses, L_{ED} , L_{JH} , and L_{HC} , with the hyperparameters of λ_1 , λ_2 , and λ_3 , respectively. We explore a different combination of hyperparameters for loss functions and found that the whole network performs better with a combination of 0.5, 0.5, and 0.1 for λ_1 , λ_2 , and λ_3 , respectively. This combination ensures that the Joint Regressor in the Rest Network equally prioritizes the absolute joint locations estimation and maintains the joint hierarchy, and Height Classifier pays attention to height variation. We also train the Posture Classifier using *Adam* optimizer with a learning rate of 1×10^{-4} and a batch size of 4 for training epochs of 1,000. The Rest network is implemented in Python with TensorFlow 2.4 [80] using Spyder IDE [81] and Anaconda version 4.10.3 distribution [82]. Our training time varies between 6 to 10 hours for completion in a GPU server with two NVIDIA RTX A6000 nodes [83].

5 MISLEEP EVALUATION

Evaluation Summary: We summarize *MiSleep*'s overall performance below:

- *MiSleep*'s HMM-Viterbi detects toss-turn events with a median accuracy above 85% along with median precision, recall, and F1-score of 0.97, 0.88, and 0.92, respectively, indicating high accuracy and low false positives. *MiSleep* always predicts the event duration within 1.7 s of the ground truth and can identify the start and end times within 0.25 s and 0.73 s errors in the median.
- *MiSleep*'s Rest Network predicts the 3D location of body joints of a person in various sleep postures with median and 90th percentile errors of 1.3 cm and 6.24 cm, respectively. The network generalizes well across multiple volunteers, with median and 90th percentile errors of 2.3 cm and 7.4 cm, respectively. The network shows similar performance for persons sleeping under blankets.
- The Rest Network classifies volunteers based on their height information with accuracies ranging from 87% to 94% for eight volunteers and classifies five different postures with an accuracy above 90% for all eight volunteers and generalizes well across unseen volunteers.

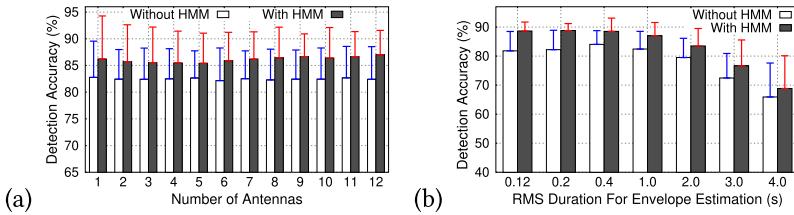


Fig. 13. Rest and toss-turn detection accuracy: Effects of (a) the number of antennas. (b) different RMS duration for envelope estimation. The bar and errorbar represent the median and standard deviation across 19,386 states. We have three observations: (1) Performance is consistently better with HMM-Viterbi. (2) Only a small improvement is observed when using multiple antennas. (3) A large RMS duration causes more state prediction error, since the reaction time of the system becomes slower.

5.1 Toss-turn Detection Results

5.1.1 State Detection Performance. To evaluate the effectiveness of *MiSleep*'s toss-turn detection modules, we use the dataset collected from a single volunteer performing several toss-turns on the bed. *First*, we preprocess the dataset following the method in Section 3.2 to generate the input-output pairs of mmWave reflected signals and 3D location of body joints. Also, for each pair, we collect the corresponding Kinect depth images to identify the ground truth rest or toss-turn states. *Then*, we build the HMM from the Kinect ground truth, use the reflected signals to estimate the envelope, and apply the Viterbi decoder to it to predict the states. We find the ground truth toss-turn using these three steps: (1) apply a fixed mask corresponding to the area of the bed in the depth image; (2) calculate the pixel-to-pixel difference in the successive depth images; and (3) find the energy in the residual depth, which ensures that the effect of posture change is sharply amplified in the output (see Figures 6 and 7). Since it is hard to control the number of toss-turns during the actual sleep and obtain a reasonably sized dataset, our volunteer mimics the toss-turn events with posture changes within a short period of 20 to 30 seconds to generate one dataset. We obtain 19,386 state observations from 30 datasets and compare the state detection accuracy by calculating the percentage of the cases where *MiSleep*'s observation matches with the Kinect output. Since COTS mmWave devices can have a variable number of antennas, we also use the observation from multiple antennas and take median votes to decide between the output binary states.

Figure 13(a) shows the state detection accuracy w.r.t. the Kinect-based ground truth across the different number of antennas. Clearly, *MiSleep*'s HMM-Viterbi performs consistently better than the envelope thresholding algorithm, and the median accuracy is always above 85%, reaching up to 100% in certain cases. This is because the HMM-Viterbi can enforce the envelope to follow the Kinect-based toss-turn events with its state and emission matrices. Moreover, the detection accuracy is unaffected by the number of antennas used in our model. This is because mmWave signals are sensitive, and even one antenna, with a large beamwidth, can cover the whole bed area and detect state changes. Furthermore, we use a variable number of frames, from 3 to 100, corresponding to 0.12 to 4 s, to compute the RMS for the envelope detector and predict the states.

Figure 13(b) shows that the large RMS duration, such as 4 s, although useful for suppressing false detections, decreases the state detection accuracy significantly, since it reacts slowly to the true state changes. Still, *MiSleep* performs consistently better with HMM-Viterbi, and RMS duration of 1 s shows 88% detection accuracy on the median. Figure 14(a) shows the distribution of precision, recall, and F1-score of the event detection, where the median values are 0.97, 0.88, and 0.92, respectively, indicating *MiSleep* is not only accurate but also has low false-detection rates. Finally, these state detection performances are estimated during a short period by limiting our observations near

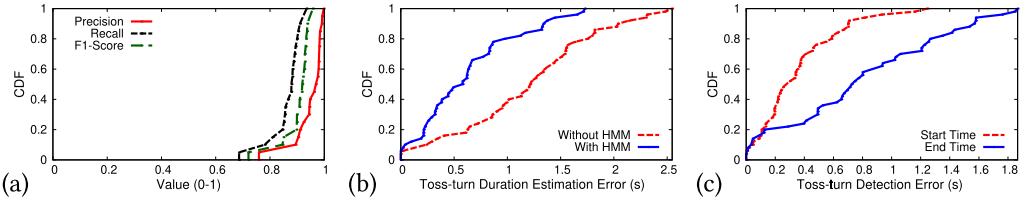


Fig. 14. (a)Precision, recall, and F1-score for the state detection with HMM. (b) Distribution of errors in toss-turn duration estimation, with and without HMM. (c) Distribution of errors in toss-turn start and end times detection with HMM.

the toss-turn events. When we test the reflected signals from only the static postures, collected over a combined duration of more than 7.5 hours, we observe that *MiSleep* always outputs the rest states, i.e., it has 100% accuracy, irrespective of the volunteers or their postures. This is because mmWave signals have very low fluctuations under the static posture and are unaffected by the quiet bedroom conditions. *In summary, these results show that MiSleep’s toss-turn detector modules consistently perform well and can predict the rest and toss-turn states with high accuracy.*

5.1.2 Toss-turn Timing Parameters. We also evaluate *MiSleep*’s performance in identifying the timing of the toss-turn events. This information could be useful in not only identifying the precise start and end of toss-turn but also annotating the events automatically. To this end, we use the same set of state observations as before and estimate the toss-turn times from both *MiSleep* and ground truth. We evaluate three different errors in timing parameters: toss-turn start time, end time, and duration. For the start and end times, we first locate each event in the ground truth and identify the time of the state change from 0 to 1 (i.e., rest to toss-turn) as start and 1 to 0 (i.e., toss-turn to rest) as an end. For each case, we identify the closest time of such events detected by *MiSleep* and estimate their corresponding start and end times. For the duration error, we find the sum of the absolute differences in start and end times of the ground truth and *MiSleep*.

Figure 14(b) shows the distribution of error in duration estimation across ~ 100 toss-turn events and compares the performance with and without the HMM-Viterbi. *First*, our total count for predicted toss-turn events shows that *MiSleep* did not miss any detection. *Second*, without HMM-Viterbi, we observe that the median and 90th percentile errors are 1.22 s and 2.04 s, respectively. In contrast, HMM-Viterbi can reduce this error to 0.58 s and 1.34 s in median and 90th percentile, respectively. More importantly, *MiSleep* always predicts the duration within 1.7 s of the ground truth across all our observed events. Figure 14(c) further shows the toss-turn start and end time estimation errors with HMM-Viterbi. Here, the median errors in start and end detections are 0.25 s and 0.73 s, respectively. Moreover, the 90th percentile errors show that *MiSleep* is accurate within 1 s and 1.5 s to detect the start and end of the event, respectively. *These experimental results show that MiSleep not only can count the number of toss-turn events accurately but also is highly accurate in locating the events in time.*

5.2 Sleep Posture Prediction Results

5.2.1 Error in Estimating 3D Location of Body Joints. We now evaluate the performance of *MiSleep*’s Rest Network in predicting the 3D location of body joints during sleep. For a baseline performance, we first use a small-scale dataset of $\sim 9,730$ samples collected from three volunteers (two females and one male, height varying from 155 cm to 178 cm), with the lowest and highest height among all volunteers, performing five different sleep postures and their variations, and then evaluate the performance across all eight volunteers. Initially, we want to evaluate the performance of the system on volunteers with the lowest and highest height among all volunteers, such that,

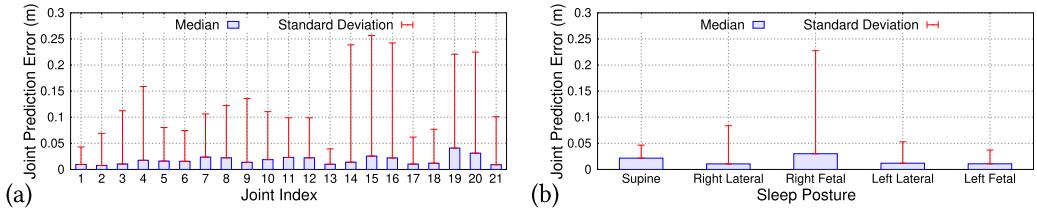


Fig. 15. (a) *MiSleep*'s prediction errors for 3D locations of 21 joints across five poses for three volunteers. (b) *MiSleep*'s prediction errors across all joints, summarized for individual sleep postures.

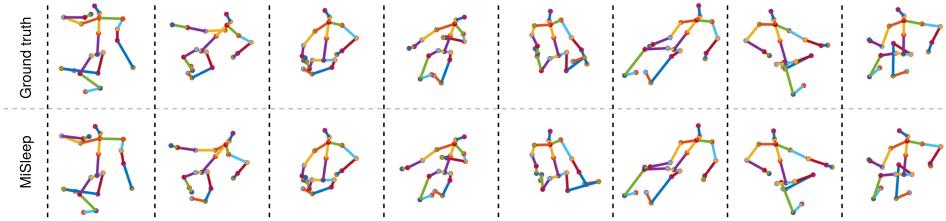


Fig. 16. Various sleep postures: (top) ground truth; (bottom) predicted.

when we evaluate the generalizability of the system, we can test on unseen volunteers whose height falls within the range. We follow the same process as before to synchronize and resample the ground truth joint locations and mmWave reflected signals, and we label the volunteers' height into discrete categories following Section 4. Then, we randomly select $\sim 8,700$ samples for training and $\sim 1,030$ samples for testing. All our samples are evenly distributed across all postures. During training, we use 20% of the training samples for validation. The baseline results include the performance of the Joint Regressor in terms of the Euclidean distance between the ground truth and predicted joint locations.

Figure 15(a) shows the performance of *MiSleep*, where we observe that for all 21 joints, median error is always less than 4.1 cm. However, we see a very high standard deviation across joints 14, 15, 16, 19, and 20 (left knee, left ankle, left foot, right ankle, and right foot). This increased standard deviation can be attributed to limitations in the device's angular resolution. If the angular resolution is not sufficiently fine, then two body parts that are close together might be detected as one, or their relative positions might be inaccurately assessed. While our kinematics-based learning model offers robustness against such challenges, it is not entirely impervious to them. To investigate this issue further, we plot the aggregated errors from all joints and separate them in terms of the sleep postures. Figure 15(b) shows that a majority of the errors are from the right fetal, i.e., a curled up posture. The reason for such high error could be due to the inability of the ground truth device to produce accurate joint locations for curled-up postures. Although the location of feet joints play a significant role in sleep postures. But the joints that are critical to facilitate a sleep posture monitoring application can be predicted accurately by *MiSleep*. Figure 16 shows top-view of skeletons for various sleep postures predicted by *MiSleep*. These results demonstrate that *MiSleep* can predict the 3D location of body joints accurately, and the predicted postures are similar to the ground truth.

5.2.2 Effect of Height Classifier. Recall that *MiSleep* uses the output from the Height Classifier to fine-tune its Joint Regressor so the network learns association between height and the 3D posture of a person from the mmWave signals. Our model's primary objective is to accurately deduce the

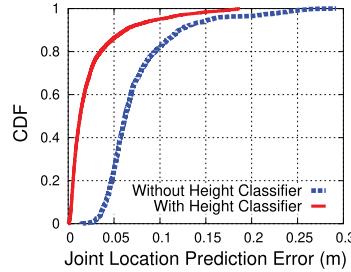


Fig. 17. Height classifier improves the performance of Rest Network. The median joint location error is reduced by 4.92 cm.

3D joint positions, and these positions remain relatively consistent across different body types when considering individuals of similar height. If two people of the same body type and similar height were to assume the same sleeping posture, then our system would generate nearly identical skeletons based on the reflected signals. This improves the generalization ability of the network as well as helps to refine the prediction.

To understand the benefit of the Height Classifier, we estimate the absolute 3D joint location errors with and without using it in the model. To this end, we first train the Rest Network without the classifier on $\sim 6,500$ samples collected from three volunteers and test it on another set of $\sim 2,000$ samples. Furthermore, we build the Height Classifier following Section 3.4.2 into the network and feed its loss function to fine-tune the output of the Joint Regressor. Then, we evaluate the performance with the same set of training and testing samples.

Figure 17 shows the performance of the Rest Network with and without the Height Classifier. We observe that *MiSleep* predicts joint locations with median and 90th percentile errors of 6.22 cm and 12.12 cm, respectively, without the Height Classifier. However, by incorporating the classifier, we observe a better prediction with median and 90th percentile errors of 1.3 cm and 6.24 cm, respectively. This is because the network can better associate height information of an individual with variations in sleep postures, which, in turn, enables better joint location estimation from the reflected mmWave signals. *These results show the efficacy of the Height Classifier in assisting MiSleep’s Rest Network to yield better estimation of the 3D location of body joints from mmWave reflected signals.*

5.2.3 Effect of Number of Volunteers. To evaluate the generalizability of *MiSleep*’s Rest Network for diverse volunteers, we now perform an ablation study. Here, we would like to understand the performance and amount of fine-tuning required for new, unseen volunteers for *MiSleep*. To this end, we randomly select 2,000 test samples from eight volunteers, including all five sleep postures, with 250 samples from each volunteer. These are unseen data for *MiSleep*’s Rest Network. We then create a training set of $\sim 3,000$ samples from one volunteer and train *MiSleep*’s Rest Network: We consider it as a base model. We then evaluate the performance on the test samples that include data from all volunteers by calculating absolute joint location error across all 21 joints and the absolute error in height prediction. Then, we progressively add two new volunteers’ datasets and fine-tune the base model and test on the same set of test samples for eight volunteers.

Figure 18 shows the performance of *MiSleep* with different levels of fine-tuning. With zero additional volunteers for the base model, the network is unable to capture variations in sleep posture and its relation to the height of varying individuals. We see that the median joint location error is very high, 11.6 cm, and the predicted body joints may not be usable in practice. Similarly, the median error in the predicted height of the unseen volunteers could be 10.2 cm, which is highly

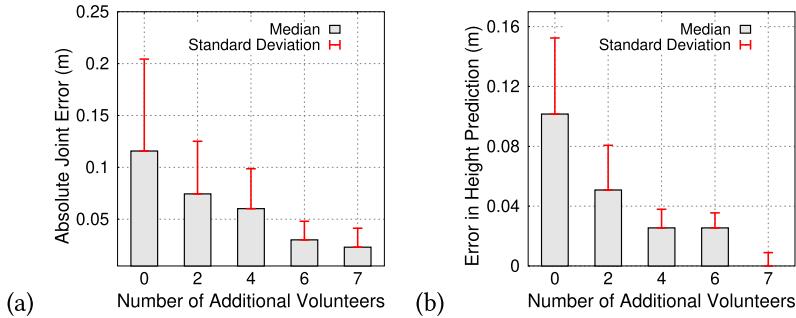


Fig. 18. *MiSleep* generalizes across multiple volunteers with only a little fine-tuning. (a) Joint and (b) Height errors. *MiSleep* starts with training from a single volunteer’s dataset, and as we progressively add new volunteers, the error decreases for both joint locations and height estimation. The testing samples are from all eight volunteers.

Table 4. Confusion Matrix of Three Volunteers Classification in *MiSleep*

Actual/Predicted	Volunteer 1	Volunteer 2	Volunteer 3
Volunteer 1	93.6%	2.9%	3.5%
Volunteer 2	4.7%	92.7%	2.6%
Volunteer 3	3.8%	1.8%	94.4%

inaccurate. This is intuitive, since the network has learned from the dataset of only one volunteer, which results in both body joint and absolute height errors. However, by fine-tuning the network with two additional volunteers’ datasets for 500 epochs, we see an improvement in prediction as median errors for joint locations and height decrease to 7.5 cm and 5.08 cm, respectively. This is because *MiSleep* can learn feature association between individuals and their sleep posture and become more generalizable to capture the correlation between mmWave reflections and human body shape. Such improvements are also consistent in both the joint locations and height prediction, as we increase the number of volunteers for fine-tuning.

In summary, MiSleep is generalizable to different volunteers and requires minimal fine-tuning as it learns an association between individual’s skeletal feature and reflections to predict posture.

5.2.4 Classification of Different Individuals. Recall that *MiSleep*’s Rest Network also provides the height information from its Height Classifier, which is used to further fine-tune the Joint Regressor’s performance. To evaluate the efficacy of the classifier, we use the datasets from three volunteers with different heights and label them into three different classes and predict the labels from the model. We randomly test on 1,026 data samples (342 from each volunteer), and Table 4 shows the confusion matrix of categorical labeling, where columns are the actual labels and rows are the predicted labels. While the misclassification rate between volunteers 1 and 2 is a bit higher because they have similar heights (4.3% for 1 and 2 vs. less than 2% for the others), we see an average classification accuracy of 93.56% across all three classes. Furthermore, Table 5 shows the confusion matrix of all eight volunteers’ height prediction, where we retrain *MiSleep* with 24,000 data samples that include data from all volunteers. Then, we randomly test on 2,000 data samples (250 per volunteer) and predict their height labels. We observe that class prediction is consistent with an accuracy of more than 85% across all volunteers. We also observe some misclassification rates of 0.4% to 8% among volunteers that share nearly the same height information. Even though the error is a bit higher, the overall performance of the classifier is acceptable with accuracy in

Table 5. Confusion Matrix of Eight Volunteers Classification in *MiSleep*

Actual/Predicted	Volunteer 1	Volunteer 2	Volunteer 3	Volunteer 4	Volunteer 5	Volunteer 6	Volunteer 7	Volunteer 8
Volunteer 1	91.2%	4.8%	2%	2%	0%	0%	0%	0%
Volunteer 2	6.4%	88.8%	3.2%	1.6%	0%	0%	0%	0%
Volunteer 3	4.4%	1.6%	87.2%	6.4%	0.4%	0%	0%	0%
Volunteer 4	0%	0%	8%	92%	0%	0%	0%	0%
Volunteer 5	0%	0%	4%	0%	93.6%	0%	2%	0.4%
Volunteer 6	0%	0%	0%	0%	7.2%	88.8%	3.2%	0.8%
Volunteer 7	0%	0%	0%	0%	0%	92%	8%	
Volunteer 8	0%	0%	0%	0%	0%	0%	6%	94%

Table 6. Confusion Matrix of Posture Classification Trained with Data from Eight Volunteers and Tested on Those Eight Volunteers

Actual/Predicted	Supine	Right Lateral	Right Fetal	Left Lateral	Left Fetal
Supine	91.3%	4.2%	3.3%	0.9%	0.3%
Right Lateral	0%	98%	2%	0%	0%
Right Fetal	0%	2%	98%	0%	0%
Left Lateral	0%	0%	0%	98.6%	1.4%
Left Fetal	0%	0%	0%	1%	99%

Table 7. Confusion Matrix when Trained with Data from Six Volunteers but Tested on Eight Volunteers

Actual/Predicted	Supine	Right Lateral	Right Fetal	Left Lateral	Left Fetal
Supine	90.7%	4.8%	3.8%	0.5%	0.2%
Right Lateral	0%	97.4%	2.6%	0%	0%
Right Fetal	0%	6.3%	93.7%	0%	0%
Left Lateral	0%	0%	0%	98.4%	1.6%
Left Fetal	0%	0%	0%	2.2%	97.8%

the range of 87% to 94%. These results demonstrate that *MiSleep* can classify different volunteers accurately and can generalize well to individuals of varying heights and genders.

5.2.5 Classification of Different Sleep Postures. We now evaluate *MiSleep*'s ability to predict the discrete categories of broad sleep postures (Figure 11). Such broad categories could be used to not only label the rest states but also summarize an individual's sleeping habits. We use the datasets from all volunteers performing the five postures and train the Posture Classifier using 7,336 training samples (917 from each volunteer) and randomly select 3,735 test samples (747 from each posture class) from all volunteers to evaluate the classification accuracy. Table 6 shows the confusion matrix of categorical labeling with columns as an actual label and rows as predicted label for sleep postures. We observe that the accuracy for the broad categories of posture is well above 91%, which indicates that *MiSleep* can accurately extract distinguishing features from postures by identifying signatures in reflected signals associated with each posture.

Furthermore, to observe the generalizability of the Posture Classifier, we reduce training samples to ~5,500 by eliminating samples from two volunteers. Then, we predict using the same 3,735 test samples from all volunteers (747 from each of the categories). Table 7 shows that even by training *MiSleep* with datasets from six volunteers and testing on datasets from eight volunteers, there is no significant difference in the classification performance. From these results, we see that the mmWave signal carries a distinct signature associated with a person's sleep posture, and *MiSleep* can classify these postures with very high accuracy.

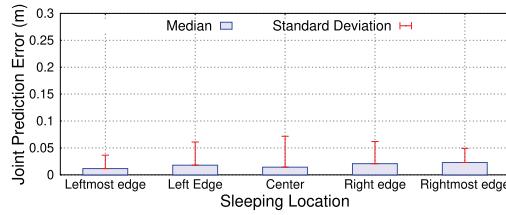


Fig. 19. Average prediction error for body joint location for five sleeping locations in bed.

5.2.6 Effect of Different Sleeping Positions on the Bed. We now evaluate the effect of body's location in bed by asking a volunteer to sleep in five different locations in the bed. Since the location of a person may change during the sleep, we would like to verify if *MiSleep* would still work without re-training the system. Intuitively, *MiSleep* should perform similarly in all locations when the background is static. It is because *MiSleep* relies on the signatures in reflected signals for specific postures. Therefore, if there is no significant change in background, *MiSleep* generates similar reflected signals for the same postures at every location. We select five different locations in bed: (1) leftmost edge (90 cm from bed center), (2) left edge (45 cm from bed center), (3) center, (4) right edge (45 cm from bed center), and (5) rightmost edge (90 cm from bed center). Figure 19 shows the prediction for these five locations. We observe a median error ranging from 1.17 to 2.29 cm, which is on par with the baseline prediction. Our observation shows bed location has minimal impact on the performance of *MiSleep* in predicting joint locations. *Therefore, MiSleep has a similar prediction performance regardless of the body's locations in the bed.*

5.2.7 Effect of Occlusion. In all our analyses thus far, we have evaluated sleep postures without any occlusion, such as a blanket. We now evaluate the effect of occlusion on *MiSleep*'s performance. We rely on Kinect to obtain the ground truth sleep postures, but Kinect does not work under the blanket, since its depth sensors cannot penetrate through occlusion. Therefore, to evaluate the performance under the blanket, we rely on approximating the ground truth. First, we ask a volunteer to perform sleep posture without a blanket so Kinect can capture the ground truth. Then, we cover their body with a blanket without disturbing the setup, ensuring no change in the sleep posture and collecting the mmWave reflections. From the cross-correlation between consecutive reflected signals (Figure 20(a)), we ensure the mmWave reflections are not used for prediction when we cover the person with the blanket. Such approximation works, since the signature of the reflected signal for the same posture does not change drastically with or without the blanket: mmWave signal can still penetrate through the blanket (see Figure 20(b)). Also, we collect data with two types of blanket: thin sheet and thick comforter, as shown in Figure 21. In total, we collect 600 data samples for each case with the volunteer in various sleeping postures. We compare *MiSleep*'s prediction for occlusion to the ground truth obtained from Kinect before the occlusion.

Our experimental results in Figure 20(c) show median and 90th percentile errors of 3.5 cm and 4.45 cm, respectively, when a person performs posture without the blanket, which is closer to our baseline performance. With the thin blanket, the median and 90th percentile errors are 6.8 cm and 9.2 cm, respectively, and with the thick blanket, they are 11 cm and 13.33 cm, respectively. Note that we do not fine-tune our system due to a lack of ground truth. Our visual result in Figure 21(a–c) shows the prediction of *MiSleep* without the blanket, under the thin blanket, and under the thick blanket. Even without fine-tuning, we observe that *MiSleep* can capture the sleep postures during occlusion. We believe with better medium to capture the ground truth, such as a pressure sensor mattress [84], we can improve the performance of *MiSleep* to capture postures under the blanket.

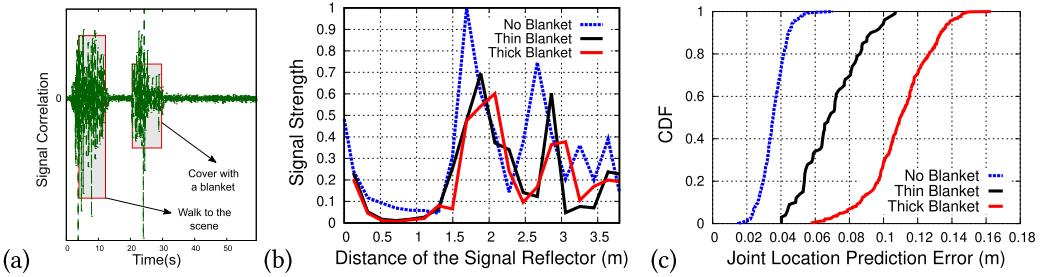


Fig. 20. (a) Cross-correlation of reflected signals during data collection. (b) Reflected signal strength with no blanket, thin blanket, and thick blanket. (c) CDF plots for prediction error from all three settings.

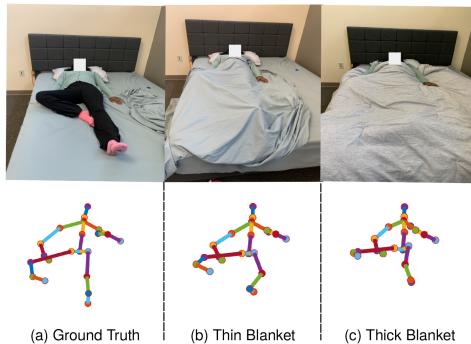


Fig. 21. (Top) RGB images of sleep postures. (Bottom) Predicted sleep postures for three settings: (a) Without blanket; (b) With thin blanket; (c) With thick blanket. Sleep postures are from a 2D (top) view.

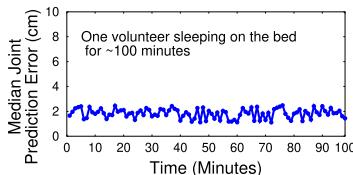


Fig. 22. Median prediction error of 3D locations of body joints for ~ 100 minutes of sleep.

5.2.8 Longitudinal Study. We now demonstrate *MiSleep*'s Rest Network performance for a longer duration. We collect 98 instances of various postures of a volunteer, each posture spanning 60 seconds to create an observational window of 98 minutes. Then, we compute the median prediction error of body joint locations for each instance. Figure 22 shows that even for several variations of sleep postures, our median error lies within a tolerable range of 1.1 to 2.5 cm, indicating the effectiveness of Rest Network in capturing the variations of sleep postures for a long duration. *The result shows that the prediction performance of MiSleep is unaffected by the observational period, so it can identify the sleep postures throughout the night with tolerable errors.*

6 COMPARISON WITH A STATE-OF-THE-ART WORK

Since our proposed approach is the first of its kind in using mmWave wireless signals for 3D joint estimation during sleep, our evaluation is limited to comparing it with other works that concentrate on activities such as standing or performing exercises. Specifically, we compare *MiSleep* with

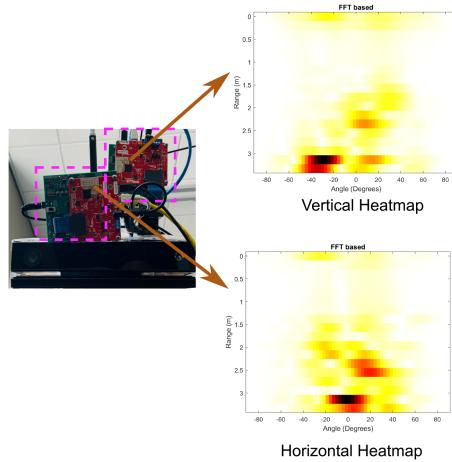


Fig. 23. Heatmaps generated from horizontal and vertical mmWave devices.

RF-Pose [85], a state-of-the-art technique in the RF domain that estimates human poses by generating skeletons from radio signals during standing postures.

In *RF-Pose*, a student-teacher network approach is implemented, where the teacher network learns from RGB images and demonstrates high accuracy in estimating human pose. In our system, our ground truth pose information from Kinect's depth images is highly accurate and, thus, acts as the output of teacher's network to guide the student network. Similar to *RF-Pose*, the network learns to estimate the pose by computing and comparing the loss between the output of the teacher network (based on Kinect's human pose) and the student network's predictions. Based on the loss, the network adjusts its parameters to minimize the difference between output of teacher network and student network's prediction during training. As the training progresses, the student network gradually improves its performance and aligns its pose estimation with the output of teacher. This iterative process of minimizing the loss helps the student network refine its pose-estimation capabilities and achieve higher accuracy over time.

We follow similar architecture as in *RF-Pose*, making some adjustments to match our specific input and output requirements. Like in *RF-Pose*, our student network follows an encoder-decoder design, but with modifications to accommodate the input from our horizontal and vertical mmWave devices, as shown in Figure 23. Our heatmaps are generated from the horizontal and vertical antenna array of our mmWave devices. The size of these heatmaps is 26×64 . Since sleep is a static activity, we do not consider the temporal dimension and set it to 1. Consequently, our input to the encoder is represented as $1 \times 26 \times 64 \times 1$, indicating the number of channels \times depth \times height \times width. This modified encoder-decoder architecture allows us to process the heatmap information from our mmWave devices effectively and estimate human pose accurately for sleep monitoring.

Similar to the architecture used in *RF-Pose*, our encoder consists of 10 layers of spatio-temporal convolutions with a filter size of $9 \times 5 \times 5$. Every other layer applies $1 \times 2 \times 2$ strides on spatial dimensions. Batch normalization and ReLU activation functions are applied after each layer, and max-pooling is performed every second layer. The encoder takes the horizontal and vertical heatmaps as input and encodes them into a $32 \times 2 \times 1 \times 1$ tensor, separately for each heatmap. For the decoder, we utilize a combination of spatio-temporal convolutions and fractionally strided convolutions to decode the pose into a 21×3 tensor. The overall architecture is similar to *RF-Pose*, but we modify the output layer to generate the 21×3 pose information. To train the network, we use the mean-squared error as the loss function, which measures the difference between the ground

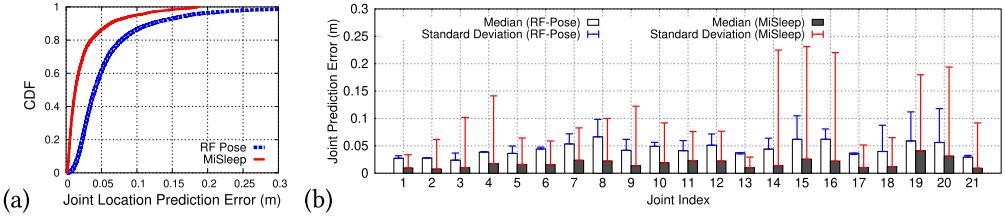


Fig. 24. (a) Performance comparison of joint location prediction between *MiSleep* and *RF-Pose*. (b) *RF-Pose*'s prediction errors for 3D locations of 21 joints.

truth pose and the generated pose. The network is trained using a batch size of 24, with the Adam optimizer and a learning rate of 0.001. The implementation is done in the PyTorch framework.

We compare the performance of *MiSleep*'s Rest Network and *RF-Pose* in terms of the Euclidean distance between the ground truth and predicted joint locations in Figure 24. In Figure 24(a), we observe that *RF-Pose* predicts joints with median and 90th percentile errors of 4.04 cm and 12.14 cm, respectively, while our baseline result shows median and 90th percentile errors of 1.3 cm and 6.24 cm. Additionally, we analyze the prediction error for individual joints in Figure 24(b). We observe that for all 21 joints, the median error is always less than 6 cm, which is comparable to our baseline in Figure 15(a), where the median error is always less than 4 cm across all 21 joints. However, we notice that with *RF-Pose*, the standard deviation across all joints consistently reduces, indicating lower variability across data points.

*We observe that implementing *RF-Pose*'s architecture for our input yields results that are quite similar to our baseline, indicating that the joint predictions produced by *MiSleep* are comparable to those achieved by the state-of-the-art human joint prediction system.*

7 LIMITATIONS AND FUTURE WORKS

Implementing *MiSleep* on mmWave Networking Devices: We have designed and prototyped *MiSleep* on COTS mmWave Radars that operate on a 1.62 GHz bandwidth and can achieve a depth resolution of 9.26 cm. In the future, our goal is to deploy *MiSleep* for at-home monitoring by reusing COTS mmWave networking devices, such as the 5G home wireless routers. We expect that *MiSleep* will perform even better on COTS mmWave communication devices. This is because current standard mmWave networking devices, which follow IEEE 802.11ad protocol [48], operate on a 2 GHz bandwidth; so it can achieve a better depth resolution of 7.5 cm. With higher depth resolution, a device can distinguish reflections from two different points with higher confidence, and it should be able to predict joint locations with higher accuracy. Also, these devices can already measure the reflected signals from the **Channel Estimation (CE)** header field [48], requiring no protocol or packet format modifications. But current COTS mmWave networking devices do not provide user access to the reflected signals yet without device firmware modifications. We propose to explore its feasibility in the future.

Extending *MiSleep* for Two Persons on a Bed: *MiSleep* is currently designed and evaluated as a single person sleep posture monitoring system. Extending *MiSleep* to monitor two persons simultaneously is very challenging. *First*, when two persons share a bed, they lie in proximity to each other, so, the reflected signals from their bodies will be entangled with each other when they are received by *MiSleep*. Given the depth resolution is 9.26 cm only, it might be hard to separate the mixed reflections. *Second*, two persons may move and toss-turn at the same time, which makes it harder to identify and track the toss-turn of an individual based on signal cross-correlation. One approach could be to track vital signs, such as breathing rate, heart rate, and so on, as two different

individuals may show some distinct signatures, and then, we can improve spatial resolution with digital beamforming overall receive antennas. Existing works [86–89] can extract vital signs for multi-person or improve spatial resolution to enable multi-person tracking. But they are limited to human motion and have not been adapted for sleep posture monitoring, and we will investigate their feasibility and challenges in the future.

Long-term Sleep Monitoring and Field Deployment: In *MiSleep*, we analyze two critical aspects of the sleep posture monitoring: (1) we identify the toss-turn events during sleep from reflected mmWave signals and classify the sleeping period to binary states; (2) we show that mmWave signals carry enough information and build and validate a model to predict the accurate 3D location of body joints. However, our observations and evaluations are limited to datasets collected for a short duration due to low buffer sizes in mmWave devices and Kinect image acquisition [90], and they quickly fill up with high sampling rates (25 fps and 30 fps). Therefore, we were unable to collect data for a long duration and analyze the joint performance of *MiSleep*'s toss-turn detector and sleep posture predictor. One option could be lowering the sample rates to 1 or 2 fps and/or modifying data collection software to enable real-time frame streaming to either a local PC or cloud. In the future, we will extend the data collection period for long-term monitoring and evaluate *MiSleep*'s performance end-to-end.

8 RELATED WORK

RF-based Sleep-posture Detection: Significant research efforts have been directed to understanding and inferring sleep posture using RF signals. These approaches use signals that bounce off the body of the person to determine the sleep posture. Reference [91] embeds RFID tags in bedsheets and enables a contactless sleep-monitoring system by recognizing sleep postures using a machine learning classifier. Reference [92] implements passive RFID tags under a bed sheet to provide a non-intrusive and comfortable way of monitoring the sleep postures. Reference [39] uses signals from a Wi-Fi like device to identify the angular orientation of a person to determine sleep posture. Reference [40] also uses Wi-Fi to exploit the fine-grained channel information to capture the minute movements caused by breathing and heartbeats. References [41, 93, 94] adopt off-the-shelf commodity Wi-Fi devices to monitor sleep, but they mainly focus on predicting sleep stages and identifying breathing and motion patterns rather than sleep postures. However, all these works provide a coarse representation of the sleep and cannot identify the key body joint locations due to the low resolution offered by Wi-Fi-like devices. MmWave signals in ubiquitous commodity networking devices can enable such system by representing the human body at a fine-grain scale as compared to Wi-Fi. Yet, it is challenging to extract body joint information directly from traditional mmWave imaging during sleep.

Millimeter-wave-based 3D Joint Estimation: In recent years, researchers have been able to extract meaningful information about humans using mmWave wireless signals from commodity devices [95–97]. In particular, extracting skeletal information has been the main focus, as it provides visual information about the 3D pose of a person [95, 98–103]. For example, Reference [98] predicts the 3D location of 15 key body joints during human walking using a point-cloud representation of mmWave reflection points, and it can enable application in pedestrian detection for traffic monitoring. Similarly, CLGNet estimates 3D location for 21 joints using a commodity mmWave device, where it uses a global spatial information encoding module to improve the detection performance of key points in minor body parts (such as hands or feet) and achieves an average accuracy of 0.29 m [100]. Furthermore, Reference [101] incorporates forward kinematic constraints to learn 3D skeleton without suffering from corrupted pose estimations. The input is in the form of a range-doppler spectrum that intuitively helps in encoding the range and velocity of each joint, enabling

the learning of spatial rotation angle required for incorporating kinematic constraints. All these existing works focus on human motion and have not been adapted for sleep posture monitoring. Unlike in human motion, where previous timestamp information about joints can be utilized to predict joint locations in the next frame, we cannot leverage this fact fully in sleep posture monitoring. It is because, during sleep, a person is mostly in the rest state, and changes in postures take place abruptly. *MiSleep* overcomes these challenges by designing a toss-turn detector and a static sleep posture predictor that detects a precise time of change in the posture and then predicts the postures during the rest states.

Sleep Behavior Assessment Using non-RF Modalities: Clinical approaches to monitoring sleep behavior include polysomnography, actigraphy, and ballistocardiography [104–106]. However, they require a patient to sleep overnight or for many hours with many sensors. Sleep behavior monitoring can be categorized into three aspects: sleep quality, sleep pattern, and sleep posture. In *MiSleep*, we focus on monitoring the sleep postures to provide visual information related to the 3D location of body joints during sleep. Accelerometer-based sleep posture monitoring system categorizes sleep posture based on accelerometer readings, where a person is required to wear a device and, based on the sensor reading, posture can be classified [107–109]. Reference [110] designs a textile-based sensor system that can be worn like a pajama to measure physiological signals and classify sleep postures. Pressure sensor-based approach, such as Reference [34], uses thousands of tiny stretch sensors embedded in textiles on top of the bed mattress for sleep posture classification. Reference [111] uses multimodal information from the pressure sensor and an RGB camera to classify postures. They rely on distributed pressure sensors or arrays of force-sensing resistors and measure the pressure exerted by the body during sleep [33, 112–114]. The major drawback of these systems is that they require additional hardware either attached to a person or embedded in a bed to detect postures, both of which could be cumbersome and bring sleep discomfort. In contrast, *MiSleep* enables contactless sleep posture monitoring by using the mmWave technology in ubiquitous 5G smart home devices.

9 CONCLUSION

In this work, we present *MiSleep*, a single-person sleep posture monitoring system that detects the toss-turn events during sleep and predicts the 3D location of body joints accurately. *MiSleep* designs a cross-correlation and HMM-Viterbi-based event detector and a customized deep learning model-based posture predictor to overcome the challenges of poor resolution, specularity, and aliasing problems in the COTS mmWave system. The experimental results demonstrate that *MiSleep* generalizes to multiple volunteers with little fine-tuning and works well for different sleep postures. In the future, we plan to extend *MiSleep* to monitor sleep postures for two persons on the bed and collect datasets for long durations to evaluate its end-to-end performance. We believe *MiSleep* can unlock the potential of 5G mmWave systems, such as home wireless routers, in enabling non-privacy-invasive, high-quality sleep monitoring systems for masses at home.

REFERENCES

- [1] Lisa Matricciani, Yu Sun Bin, Tea Lallukka, Erkki Kronholm, Melissa Wake, Catherine Paquet, Dorothea Dumuid, and Tim Olds. 2018. Rethinking the sleep-health link. *Sleep Health* 4, 4 (2018).
- [2] Max Hirshkowitz, Kaitlyn Whiton, Steven M. Albert, Cathy Alessi, Oliviero Bruni, Lydia DonCarlos, Nancy Hazen, John Herman, Eliot S. Katz, Leila Kheirandish-Gozal, David N. Neubauer, Anne E. O'Donnell, Maurice Ohayon, John Peever, Robert Rawding, Ramesh C. Sachdeva, Belinda Setters, Michael V. Vitiello, J. Catesby Ware, and Paula J. Adams Hillard. 2015. National Sleep Foundation's sleep time duration recommendations: Methodology and results summary. *Sleep Health* 1, 1 (2015).
- [3] Siaw Cheok Liew and Thidar Aung. 2021. Sleep deprivation and its association with diseases-A review. *Sleep Medic.* 77 (2021).

- [4] Rebecca Robbins and Azizi Seixas and Lillian Walton Masters and Nicholas Chanko and Fatou Diaby and Dorice Vieira and Girardin Jean-Louis. 2019. Sleep tracking: A systematic review of the research using commercially available technology. *Curr. Sleep Med. Rep.* 5, 3 (2019).
- [5] Office of Disease Prevention and Health Promotion. 2021. Sleep Health. Retrieved from <https://www.healthypeople.gov/2020/topics-objectives/topic/sleep-health/national-snapshot>
- [6] J. D. Koninck and P. Gagnon and S. Lallier. 1983. Sleep positions in the young adult and their relationship with the subjective quality of sleep. *Sleep* 6, 1 (1983).
- [7] Jason J. Liu and Wenyao Xu and Ming-Chun Huang and Nabil Alshurafa and Majid Sarrafzadeh and Nitin Raut and Behrooz Yadegar. 2014. Sleep posture analysis using a dense pressure sensitive bedsheets. *Pervas. Mob. Comput.* 10 (2014).
- [8] Arie Oksenberg and Donald S. Silverberg. 1998. The effect of body posture on sleep-related breathing disorders: Facts and therapeutic implications. *Sleep Medic. Rev.* 2, 3 (1998).
- [9] The SUN. 2020. People Are Arguing About What Is the Best Position to Sleep in but Which One Are You? Retrieved from <https://www.thesun.co.uk/fabulous/10791045/sleep-position-best-people-argue-night/>
- [10] Thoracic and Sleep Group Queensland. 2021. What Is the Best Position to Sleep in? Retrieved from <http://thoracicksleep.com.au/blog/what-is-the-best-position-to-sleep-in/>
- [11] Hedok Lee, Lulu Xie, Mei Yu, Hongyi Kang, Tian Feng, Rashid Deane, Jean Logan, Maiken Nedergaard, and Helene Benveniste. 2015. The effect of body posture on brain glymphatic transport. *J. Neurosci.* 35, 31 (2015).
- [12] Tom Colquitt. 1987. The sleep-wear syndrome. *J. Prosthet. Dentist.* 57, 1 (1987).
- [13] Steven J. McCabe and Anna L. Uebel and Vasyl Pihur and Roberto S. Rosales and Isam Atrosi. 2007. Epidemiologic associations of carpal tunnel syndrome and sleep position: Is there a case for causation? *HAND* 2, 3 (2007).
- [14] Laura Beltran-Agulló, Yvonne M. Buys, Farzana Jahan, Colin M. Shapiro, John G. Flanagan, Jason Cheng, and Graham E. Trope. 2017. Twenty-four hour intraocular pressure monitoring with the SENSIMED triggerfish contact lens: Effect of body posture during sleep. *Brit. J. Ophthalmol.* 101, 10 (2017).
- [15] Tamis Pin, Beverley Eldridge, and Mary P Galea. 2007. A review of the effects of sleep position, play position, and equipment use on motor development in infants. *Develop. Medic. Child Neurol.* 49, 11 (2007).
- [16] Gustavo Desouzart, Rui Matos, Filipe Melo, and Ernesto Filgueiras. 2013. Effects of sleeping position on back pain in physically active seniors: A controlled pilot study. *Work* 53, 2 (2013).
- [17] Anshika Singh and Sumit Raghav. 2015. A systematic review protocol on correlation between sleep body posture and cervical spine dysfunction among the patients with mechanical neck pain. *Int. J. Multidiscip. Educ. Res.* 8, 10 (2015).
- [18] Susan J. Gordon and Karen A. Grimmer and Patricia Trott. 2007. Sleep position, age, gender, sleep quality and waking cervico-thoracic symptoms. *Internet J. Allied Health Sci. Pract.* 5, 1 (2007).
- [19] Simon A. Joosten and Denise M. O'Driscoll and Philip J. Berger and Garun S. Hamilton. 2014. Supine position related obstructive sleep apnea in adults: Pathogenesis and treatment. *Sleep Medic. Rev.* 18, 1 (2014).
- [20] Ambra Stefani and Birgit Högl. 2020. Sleep in Parkinson's disease. *Neuropsychopharmacology* 45, 1 (2020).
- [21] Jennifer A. Liebenthal and Shasha Wu and Sandra Rose and John S. Ebersole and James X Tao . 2015. Association of prone position with sudden unexpected death in epilepsy. *Neurology* 84, 7 (2015).
- [22] Sleep Foundation. 2022. Best Sleeping Positions. Retrieved from <https://www.sleepfoundation.org/sleeping-positions>
- [23] Johns Hopkins. 2022. Choosing the Best Sleep Position. Retrieved from <https://www.hopkinsmedicine.org/health/wellness-and-prevention/choosing-the-best-sleep-position>
- [24] Mayo Clinic. 2022. Mayo Clinic Minute: What Is the Best Sleeping Position? Retrieved from <https://newsnetwork.mayoclinic.org/discussion/mayo-clinic-minute-what-is-the-best-sleeping-position/>
- [25] Orthopaedic Associates of St. Augustine. 2022. The Best Sleeping Positions for Your Neck and Spine. Retrieved from <https://www.oastaug.com/the-best-sleeping-positions-for-your-neck-and-spine/>
- [26] MFI Medical. 2022. Embla N7000 PSG System—Certified Refurbished. Retrieved from <https://mfimedical.com/products/embla-n7000-psg-system>
- [27] Po-Yuan Jeng and Li-Chun Wang and Chaur-Jong Hu and Dean Wu. 2021. A wrist sensor sleep posture monitoring system: An automatic labeling approach. *Sensors* 21, 1 (2021).
- [28] Yoshifumi Kishimoto and Akimitsu Akahori and Koji Oguri. 2006. Estimation of sleeping posture for m-health by a wearable tri-axis accelerometer. In *3rd IEEE/EMBS International Summer School on Medical Devices and Biosensors*.
- [29] Vandan Davoodnia and Ali Etemad. 2019. Identity and posture recognition in smart beds with deep multitask learning. In *IEEE International Conference on Systems, Man and Cybernetics (SMC'19)*.
- [30] Poyuan Jeng and Li-Chun Wang. 2017. An accurate, low-cost, easy-to-use sleep posture monitoring system. In *International Conference on Applied System Innovation (ICASI'17)*.
- [31] Shinjae Kwon and Hojoong Kim and Woon-Hong Yeo. 2021. Recent advances in wearable sensors and portable electronics for sleep monitoring. *iScience* 24, 5 (2021).

- [32] Amazon. 2022. TACTILUS Mattress Pressure Mapping Sensor System Tactile Force Bed Software Body. Retrieved from <https://www.amazon.com/TACTILUS-Mattress-Pressure-Mapping-Software/>
- [33] Keison Tang and Arjun Kuma and Muhammad Nadeem and Issam Maaz. 2021. CNN-based smart sleep posture recognition system. *Internet Things* 2, 1 (2021).
- [34] Georges Matar and Jean-Marc Lina and Georges Kaddoum. 2020. Artificial neural network for in-bed posture classification using bed-sheet pressure sensors. *IEEE J. Biomed. Health Inform.* 24, 1 (2020).
- [35] Lee Jaehoon and Min Hong and Sungyong Ryu. 2015. Sleep monitoring system using kinect sensor. *Int. J. Distrib. Sensor Netw.* 11 (2015). DOI : [10.1155/2015/875371](https://doi.org/10.1155/2015/875371)
- [36] Shuangjun Liu and Sarah Ostadabbas. 2017. A vision-based system for in-bed posture tracking. In *IEEE International Conference on Computer Vision Workshops (ICCVW'17)*.
- [37] Fei Deng and Jianwu Dong and Xiangyu Wang and Ying Fang and Yu Liu and Zhaofei Yu and Jing Liu and Feng Chen. 2018. Design and implementation of a noncontact sleep monitoring system using infrared cameras and motion sensor. *IEEE Trans. Instrum. Measur.* 67 (2018).
- [38] Sara M. Mohammadi and Majdi Alnowami and Sofia Khan and Derk-Jan Dijk and Adrian Hilton and Kevin Wells. 2018. Sleep posture classification using a convolutional neural network. In *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'18)*.
- [39] Shichao Yue and Yuzhe Yang and Hao Wang and Hariharan Rahul and Dina Katabi. 2020. BodyCompass: Monitoring sleep posture with wireless signals. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2 (2020).
- [40] Jian Liu and Yingying Chen and Yan Wang and Xu Chen and Jerry Cheng and Jie Yang. 2018. Monitoring vital signs and postures during sleep using WiFi signals. *IEEE Internet Things J.* 5, 3 (2018).
- [41] Feng Zhang and Chenshu Wu and Beibei Wang and Min Wu and Daniel Bugos and Hangfang Zhang and K. J. Ray Liu. 2021. SMARS: Sleep monitoring via ambient radio signals. *IEEE Trans. Mob. Comput.* 20, 1 (2021).
- [42] Moein Enayati and Marjorie Skubic and James M. Keller and Mihail Popescu and Nasibeh Zanjirani Farahani. 2018. Sleep posture classification using bed sensor data and neural networks. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.
- [43] Ting Wu and Theodore S. Rappaport and Christopher M. Collins. 2015. The human body and millimeter-wave wireless communication systems: Interactions and implications. In *IEEE International Conference on Communications (ICC'15)*.
- [44] Feng Zhang and Chenshu Wu and Beibei Wang and K. J. Ray Liu. 2021. mmEye: Super-resolution millimeter wave imaging. *IEEE Internet. Things J.* 8, 8 (2021).
- [45] Texas Instruments. 2020. IWR1443 Single-Chip 76-GHz to 81-GHz MmWave Sensor Evaluation Module. Retrieved from <https://www.ti.com/tool/IWR1443BOOST>
- [46] Gamestop. 2022. Kinect. Retrieved from <https://www.gamestop.com/gaming-accessories/controllers/xbox-one/products/microsoft-kinect-for-xbox-one/10115985.htm>
- [47] Adriano Meta and Peter Hoogeboom and Leo P. Ligthart. 2007. Signal processing for FMCW SAR. *IEEE Trans. Geoscience Remote Sens.* 45 (2007).
- [48] IEEE Standards Association. 2012. IEEE Standard for Information technology—Telecommunications and information exchange between systems—Local and metropolitan area networks—Specific requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band. *IEEE Std 802.11ad-2012 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012 and IEEE Std 802.11aa-2012)*. <https://ieeexplore.ieee.org/document/6392842>
- [49] Merrill Skolnik. 1962. *Introduction to Radar Systems*. McGraw-Hill Book Co.
- [50] Mehrdad Soumekh. 1999. *Synthetic Aperture Radar Signal Processing* (1st ed.). John Wiley & Sons, Inc.
- [51] NETGEAR, Inc. 2022. Nighthawk X10 Smart WiFi Router. Retrieved from <https://www.netgear.com/landings/ad7200/>
- [52] TP-Link Corporation Limited. 2022. Talon AD7200 Multi-Band Wi-Fi Router. Retrieved from <https://www.tp-link.com/us/home-networking/wifi-router/ad7200/>
- [53] IgniteNet. 2022. MetroLinq 10G Tri-Band Omni. (2022). Retrieved from <https://www.ignitenet.com/wireless-backhaul/ml-10g-omni/>
- [54] MikroTik. 2022. wAP 60G. Retrieved from https://mikrotik.com/product/wap_60g
- [55] Sivers Semiconductors AB. 2022. 60 GHz Evaluation Kits (EVK). Retrieved from <https://www.sivers-semiconductors.com/sivers-wireless/evaluation-kits/>
- [56] Airfide Networks. 2022. Airfide Brings High Performance Home and Enterprise 5G-NR Wireless. Retrieved from <https://airfidenet.com/>
- [57] Xinyu Zhang. 2022. M-Cube: An Open-source Programmable Millimeter-wave Experimental Platform. Retrieved from <http://m3.ucsd.edu/>

- [58] Neda Rojhani and Marco Passafiume and Matteo Lucarelli and Giovanni Collodi and Alessandro Cidronali. 2020. Assessment of compressive sensing 2x2 MIMO antenna design for millimeter-wave radar image enhancement. *Electronics* 9, 4 (2020).
- [59] Mahmoud Kazemi and Zahra Kavehvash and Mahdi Shabani. 2021. K-space analysis of aliasing in millimeter-wave imaging systems. *IEEE Trans. Microwave Theor. Techniq.* 69, 3 (2021).
- [60] Fadel Adib and Chen-Yu Hsu and Hongzi Mao and Dina Katabi and Fredo Durand. 2015. Capturing the human figure through a wall. In *ACM SIGGRAPH Asia Conference*.
- [61] Sherif Sayed Ahmed and Lorenz-Peter Schmidt. 2012. Illumination of humans in active millimeter-wave multistatic imaging. In *6th European Conference on Antennas and Propagation (EUCAP'12)*.
- [62] Yongyang Tian, Guang-He Lee, Hao He, Chen-Yu Hsu, and Dina Katabi. 2018. RF-based fall monitoring using convolutional neural networks. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3 (2018).
- [63] Xueyi Wang and Joshua Ellul and George Azzopardi. 2020. Elderly fall detection systems: A literature survey. *Front. Robot. AI* 7 (2020).
- [64] Yisak Kim and Juyoung Park and Hyungsuk Kim. 2020. Signal-processing framework for ultrasound compressed sensing data: Envelope detection and spectral analysis. *Appl. Sci.* 10, 19 (2020).
- [65] Bradley V. Vaughn. 2022. Approach to Abnormal Movements and Behaviors during Sleep. Retrieved from <https://www.uptodate.com/contents/approach-to-abnormal-movements-and-behaviors-during-sleep>
- [66] Daniel Jurafsky and James H. Martin. 2021. *Speech and Language Processing* (3rd ed.). Pearson Prentice Hall.
- [67] Tianlang Chen and Chen Fang and Xiaohui Shen and Yiheng Zhu and Zhili Chen and Jiebo Luo. 2020. Anatomy-aware 3D Human Pose Estimation with Bone-based Pose Decomposition. Retrieved from <https://arxiv.org/abs/2002.10322>
- [68] Shuang Liang and Xiao Sun and Yichen Wei. 2018. Compositional human pose regression. *Comput. Vis. Image Underst.* 176 (2018).
- [69] Martin Blazevic and Karla Brkic and Tomislav Hrkac. 2015. Towards Reversible De-identification in Video Sequences Using 3D Avatars and Steganography. Retrieved from <https://arxiv.org/abs/1510.04861>
- [70] Emre Aksan and Manuel Kaufmann and Otmar Hilliges. 2019. Structured Prediction Helps 3D Human Motion Modelling. Retrieved from <https://arxiv.org/abs/1910.09070>
- [71] Saad Albawi and Tareq Abed Mohammed and Saad Al-Zawi. 2017. Understanding of a convolutional neural network. In *International Conference on Engineering and Technology (ICET'17)*.
- [72] Alexander Toshev and Christian Szegedy. 2014. DeepPose: Human pose estimation via deep neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*.
- [73] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-scale Image Recognition. Retrieved from <https://arxiv.org/abs/1409.155>
- [74] Alan M. W. Porter. 1999. The prediction of physique from the skeleton. *Int. J. Osteoarchaeol.* 9 (1999).
- [75] Sciencing. 2018. How to Determine Height through the Skeleton. Retrieved from <https://sciencing.com/types-forensic-tests-7551951.html>
- [76] World Population Review. 2022. Average Height by State 2022. Retrieved from <https://worldpopulationreview.com/state-rankings/average-height-by-state>
- [77] Qi Wang and Yue Ma and Kun Zhao and Yingjie Tian. 2022. A comprehensive survey of loss functions in machine learning. *Annals Data Sci.* 9 (2022).
- [78] Raul Gomez. 2021. Understanding Categorical Cross-entropy Loss, Binary Cross-entropy Loss. Retrieved from https://grombru.github.io/2018/05/23/cross_entropy_loss/
- [79] Texas Instruments. 2020. DCA1000EVM: Real-time Data-capture Adapter for Radar Sensing Evaluation Module. Retrieved from <https://www.ti.com/tool/DCA1000EVM>
- [80] Open-Source. 2022. TensorFlow. Retrieved from <https://www.tensorflow.org/>
- [81] Open-Source. 2022. Spyder IDE. Retrieved from <https://www.spyder-ide.org/>
- [82] Open-Source. 2022. ANACONDA. Retrieved from <https://www.anaconda.com/>
- [83] NVIDIA. 2021. RTX A6000. Retrieved from <https://www.nvidia.com/en-us/design-visualization/rtx-a6000/>
- [84] Boditrak. Boditrak2. Retrieved from <https://www.boditrak.com/products/medical.php>
- [85] Mingmin Zhao and Tianhong Li and Mohammad Abu Alsheikh and Yonglong Tian and Hang Zhao and Antonio Torralba and Dina Katabi. 2018. Through-wall human pose estimation using radio signals. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*.
- [86] Fengyu Wang and Feng Zhang and Chenshu Wu and Beibei Wang and K. J. Ray Liu. 2021. ViMo: Multiperson vital sign monitoring using commodity millimeter-wave radio. *IEEE Internet Things J.* 8, 3 (2021).
- [87] Tianbo Gu and Zheng Fang and Zhicheng Yang and Pengfei Hu and Prasant Mohapatra. 2019. MmSense: Multiperson detection and identification via MmWave sensing. In *3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*.

- [88] Peijun Zhao and Chris Xiaoxuan Lu and Jianan Wang and Changhao Chen and Wei Wang and Niki Trigoni and Andrew Markham. 2019. mID: Tracking and identifying people with millimeter wave radar. In *15th International Conference on Distributed Computing in Sensor Systems (DCOSS'19)*.
- [89] Chenshu Wu and Feng Zhang and Beibei Wang and K. J. Ray Liu. 2020. mmTrack: Passive multi-person localization using commodity millimeter wave radio. In *IEEE Conference on Computer Communications*.
- [90] Ayushi Prajapati et al. 2018. Skeleton tracking using Microsoft Kinect for Windows on MatLab. *Int. J. Recent Res. Asp.* 5, 1 (2018).
- [91] Pei-Jarn Chen and Tian-Hao Hu and Ming-Shyan Wang. 2022. Raspberry Pi-based sleep posture recognition system using AIoT technique. *Healthcare* 10, 3 (2022).
- [92] Jia Liu and Xingyu Chen and Shigang Chen and Xiulong Liu and Yanyan Wang and Lijun Chen. 2019. TagSheet: Sleeping posture recognition with an unobtrusive passive tag matrix. In *IEEE Conference on Computer Communications (INFOCOM'19)*.
- [93] Xuefeng Liu and Jiannong Cao and Shaojie Tang and Jiaq Weni. 2014. Wi-sleep: Contactless sleep monitoring via WiFi signals. In *IEEE Real-Time Systems Symposium*.
- [94] Fan Li and Cheng Xu and Yang Liu and Yun Zhang and Zhuo Li and Kashif Sharif and Yu Wang. 2016. Mo-sleep: Unobtrusive sleep and movement monitoring via Wi-Fi signal. In *IEEE 35th International Performance Computing and Communications Conference (IPCCC'16)*.
- [95] Gawon Lee and Jihie Kim. 2022. Improving human activity recognition for sparse radar point clouds: A graph neural network model with pre-trained 3D human-joint coordinates. *Appl. Sci.* 12, 4 (2022).
- [96] Chengxi Yu and Zhezhuang Xu and Kun Yan and Ying-Ren Chien and Shih-Hau Fang and Hsiao-Chun Wu. 2022. Noninvasive human activity recognition using millimeter-wave radar. *IEEE Syst. J.* 16, 2 (2022).
- [97] Hongfei Xue and Yan Ju and Chenglin Miao and Yijiang Wang and Shiyang Wang and Aidong Zhang and Lu Su. 2021. mmMesh: Towards 3D real-time dynamic human mesh construction using millimeter-wave. In *19th Annual International Conference on Mobile Systems, Applications, and Services*.
- [98] Arindam Sengupta and Feng Jin and Renyuan Zhang and Siyan Cao. 2020. mm-Pose: Real-time human skeletal posture estimation using mmWave radars and CNNs. *IEEE Sensors J.* 20, 17 (2020).
- [99] Arindam Sengupta and Siyang Cao. 2021. mmPose-NLP: A Natural Language Processing Approach to Precise Skeletal Pose Estimation using mmWave Radars. Retrieved from <https://arxiv.org/abs/2107.10327>
- [100] Qing Wang and Kai Wang and Wa Cheni. 2020. CLGNet: A new network for human pose estimation using commodity millimeter wave radar. In *3rd International Conference on Algorithms, Computing and Artificial Intelligence*.
- [101] Wen Ding and Zhongping Cao and Jianxiong Zhang and Rihui Chen and Xuemei Guo and Guoli Wangi. 2021. Radar-based 3D human skeleton estimation by kinematic constrained learning. *IEEE Sensors J.* 21, 20 (2021).
- [102] Guangzheng Li and Ze Zhang and Hanmei Yang and Jin Pan and Dayin Chen and Ji Zhangn. 2020. Capturing human pose using mmWave radar. In *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops'20)*.
- [103] Kai Wang and Qing Wang and Feng Xue and Wa Cheni. 2020. 3D-skeleton estimation based on commodity millimeter wave radar. In *IEEE 6th International Conference on Computer and Communications (ICCC'20)*.
- [104] Laurent Giovangrandi and Omer T. Inan and Richard M. Wiard and Mozziyar Etemadi and Gregory T. A. Kovacs. 2011. Ballistocardiography—A method worth revisiting. *Annu Int. Conf. IEEE Eng. Med. Biol. Soc.* 2011 (2011), 4279–4282. DOI : [10.1109/EMBS.2011.6091062](https://doi.org/10.1109/EMBS.2011.6091062)
- [105] Sonia Ancoli-Israel, Roger Cole, Cathy Alessi, Mark Chambers, William Moorcroft, and Charles P. Pollak. 2003. The role of actigraphy in the study of sleep and circadian rhythms. *Amer. Acad. Sleep Medic.* 26, 3 (2003).
- [106] S. R. Pandi-Perumal, D. Warren Spence, and Ahmed S. BaHammam. 2014. Polysomnography: An overview. In *Primary Care Sleep Medicine*, Springer New York.
- [107] Heenam Yoon and Suhwan Hwang and Dawoon Jung and Sangho Choi and Kwangmin Joo and Jaewon Choi and Yujin Lee and Do-Un Jeong and Kwangsuk Park. 2015. Estimation of sleep posture using a patch-type accelerometer based device. In *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'15)*.
- [108] Kang-Ming Chang and Shin-Hong Liu. 2011. Wireless portable electrocardiogram and a tri-axis accelerometer implementation and application on sleep activity monitoring. *Telemed J. E. Health* 17, 3 (2011).
- [109] Xiao Sun and Li Qiu and Yibo Wu and Yeming Tang and Guohong Cao. 2017. SleepMonitor: Monitoring respiratory rate and body position during sleep using smartwatch. *Proc. ACM Interact. Mob. Wear. Ubiq. Technol.* 1, 3 (2017).
- [110] Ali Kiaghadi and Seyedeh Zohreh Homayounfar and Jeremy Gummesson and Trisha Andrew and Deepak Ganesan. 2019. Phyjama: Physiological sensing via fiber-enhanced Pyjamas. *Proc. ACM Interact. Mob. Wear. Ubiq. Technol.* 3, 3 (2019).
- [111] Weimin Huang and Aung Aung Phyo Wai and Siang Fook Foo and Jit Biswas and Chi-Chun Hsia and Koujuch Liou. 2010. Multimodal sleeping posture classification. In *20th International Conference on Pattern Recognition*.

- [112] C. C. Hsia and K. J. Liou and A. P. W. Aung and V. Foo and W. Huang and J. Biswas. 2009. Analysis and comparison of sleeping posture classification methods using pressure sensitive bed system. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.
- [113] Juha M. Kortelainen and Mark van Gils and Juha Parkka. 2012. Multichannel bed pressure sensor for sleep monitoring. In *Computing in Cardiology Conference*.
- [114] Chi-Chun Hsia and Yu-Wei Hung and Yu-Hsien Chiu and Chia-Hao Kang. 2008. Bayesian classification for bed posture detection based on kurtosis and skewness estimation. In *International Conference on e-health Networking, Applications and Services*.

Received 23 January 2023; revised 1 November 2023; accepted 19 January 2024