

Argosleep: Monitoring Sleep Posture from Commodity Millimeter-Wave Devices

Aakriti Adhikari; Sanjib Sur

Department of Computer Science and Engineering

University of South Carolina, Columbia, USA

aakriti@email.sc.edu; sur@cse.sc.edu

Abstract—We propose *Argosleep*, a millimeter-wave (mmWave) wireless sensors based sleep posture monitoring system that predicts the 3D location of body joints of a person during sleep. *Argosleep* leverages deep learning models and knowledge of human anatomical features to solve challenges with low-resolution, specularity, and aliasing in existing mmWave devices. *Argosleep* builds the model by learning the relationship between mmWave reflected signals and body postures from thousands of existing samples. Since practical sleep also involves sudden toss-turns, which could introduce errors in posture prediction, *Argosleep* 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 *Argosleep* with real data collected from COTS mmWave devices for 8 volunteers of diverse ages, gender, and height performing different sleep postures. We observe that *Argosleep* identifies the toss-turn events accurately and predicts 3D location of body joints with accuracy on par with the existing vision-based system, unlocking the potential of mmWave systems for privacy-noninvasive at-home healthcare applications.

Index Terms—Millimeter-Wave; Sleep Posture Monitoring; HMM-Viterbi; Convolutional Neural Networks.

I. INTRODUCTION

Humans spend approximately one-third of their life 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]. 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 [2]. Each of us sleeps in one of the broad categories of posture, such as supine, lateral, fetal, etc., and exhibits wide variations of them throughout the night [3]. The effect of different sleep postures has been studied widely to identify their relationship to different health conditions [4]–[6].

Specific sleep postures could 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 [7]. Infrequent turns due to impairment in control of the motor activity of Parkinson’s patients lead to parasomnia and restless leg syndrome [8]. 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 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 [9], [10]. These examples highlight the importance of a sleep posture monitoring system that can provide real spatio-temporal observations, which could help with corrections and 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 [11]. Existing at-home approaches are either cumbersome, costlier, or highly privacy-invasive [12]–[15]. Further, their performance is hindered by dark bedroom conditions and occlusion. Wireless-based solutions can overcome these challenges by inferring postures under no light without being privacy-invasive [16]–[18], but existing solutions rely on special-purpose low-frequency devices. Besides, they can only classify sleep postures into broad, discrete categories [19], 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 devices, such as access points, enabling the opportunity for bringing privacy non-invasive 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. 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 [20]. 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 *Argosleep*, 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, *Argosleep* 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, *Argosleep* 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. Furthermore, to generalize the model for diverse populations, *Argosleep* models a height classifier and uses the error in its prediction to finetune the model. We use a dataset collected from several static sleep postures from multiple volunteers, and at run-time, *Argosleep* 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. 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, *Argosleep* 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 prototype *Argosleep* with COTS devices by building a customized setup with two 77–81 GHz mmWave transceivers [21] to collect the reflected signals and a Microsoft Kinect Xbox One [22] to collect the ground truth 3D joint locations. *Argosleep* 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, *Argosleep* predicts the 3D location of body joints with a median error 1.3 cm only. Furthermore, *Argosleep* generalizes well for diverse volunteers with median and 90th percentile errors of 2.3 cm and 7.4 cm, respectively.

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, *Argosleep* 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

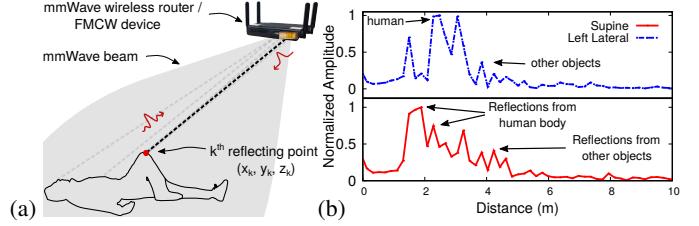


Figure 1: (a) A mmWave device captures reflected signals from the target. (b) Example of reflected signals from two sleep postures.

accelerate the research on COTS mmWave device based sleep monitoring, we will open-source our dataset and codebase.

II. BACKGROUND AND CHALLENGES

Traditional mmWave imaging approaches rely on Frequency Modulated Continuous Waves (FMCW) from a device to generate an image [23]. The device illuminates the target scene with a wideband and wide-beamwidth FMCW signal (Figure 1[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 surrounding, including the sleeping human (see Figure 1[b]). Identifying the 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 4 or 8, in the horizontal and vertical directions, [24]: 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 1024, in both horizontal and vertical directions [25], 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. (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.) Adding the reflected signals from non-uniformly spaced antennas will create spurious reflection points, and the image will appear distorted [26]. *Finally*, different body parts reflect mmWave signals differently, such as torso could reflect a strong signal, but the limbs usually reflect weak signals [20]. Such uneven reflections from different parts, 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.

III. Argosleep DESIGN

A. Overview

Argosleep 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,

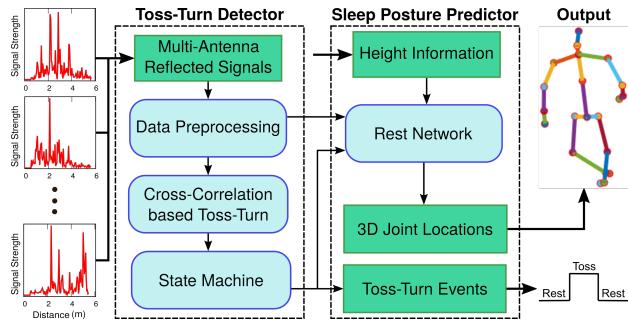


Figure 2: System overview of *Argosleep*.

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, *etc.*

To this end, *Argosleep* 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, *Argosleep* 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, *Argosleep* 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 run-time, *Argosleep* can accurately predict the joint locations only from the mmWave reflected signals. Figure 2 shows the system overview of *Argosleep*. We now describe these design components in detail.

B. Data Synchronization and Resampling

Argosleep's deep learning model relies on datasets collected from different COTS devices. Therefore, it is critical to ensure synchronization between them so that 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 allow synchronization between devices. Further, 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. Finally, processed datasets are fed into either toss-turn detector or rest network.

C. 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

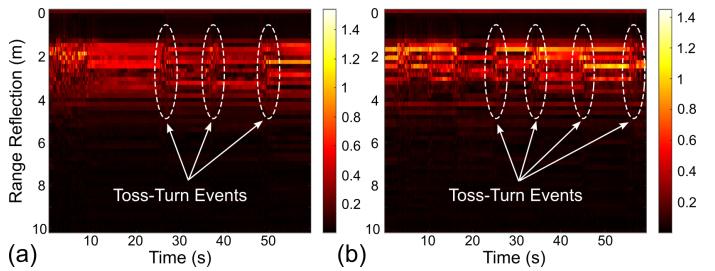


Figure 3: (a–b) STFT outputs from two monitored cases with 3 and 4 toss-turn events.

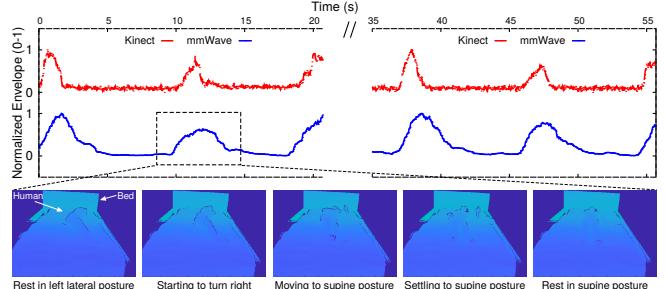


Figure 4: Monitoring toss-turn events: The envelope detector output from mmWave signals follows the changes in posture.

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 avoid erroneous predictions under toss-turns.

1) *Cross-Correlation based Toss-Turn Detection:* Inspired by the previous works on elderly fall detection using wireless signals [27], *Argosleep* 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*. To observe distinct toss-turn states, we ask a volunteer to lie down on the bed placed at 2.5 m in front of the mmWave devices for \sim 60 seconds and perform multiple toss-turns, *i.e.*, move from one posture to another. We perform a Short-Time Fourier Transform (STFT) over the signal received by one of the mmWave antennas. Figures 3(a–b) show the STFT output for two cases, with 3 and 4 toss-turn events. In comparison to fall, 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 bin of the reflected signals (Section II). So, the changes observed under the toss-turn are much weaker and do not show stark time-and-frequency changes. So, it will require additional processing to amplify the changes during the toss-turns and separate them from the rest states.

To amplify such changes, *Argosleep* 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

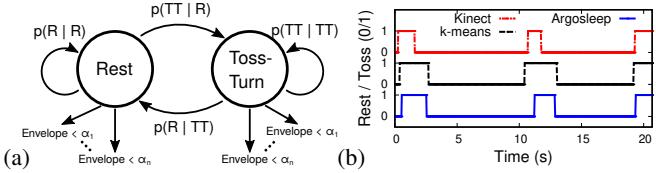


Figure 5: (a) A two-states HMM. (b) Example detection of rest and toss-turn states.

change over time should be close to zero. On the other hand, 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, *etc.*, so that the only changes due to the body movement are amplified and stand out. Let's 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, $xCorr$ can be expressed as $xCorr_t = \max_{m \in (0, n-1)} |\sum_{i=1}^{n-m} R_t(d_{i+m}) \cdot R_{t-1}^*(d_i)|$ and its time-derivative, $\Delta xCorr$ as $\Delta xCorr_{t-1}^t = xCorr_t - xCorr_{t-1}$, where R_{t-1}^* is the complex conjugate of the received reflected signal at time instant $t-1$. To reduce the number of oscillations between the false detections (+/-) and states, *Argosleep* smoothens $\Delta xCorr$ over time with an envelope detector using a Hilbert Transformation, similar to [28]. 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. On the other hand, 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.

Figure 4 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. However, using the envelope output for state detection is challenging as the output of the envelope detector is a real number between 0 to 1, where 0 indicates no change in the successive reflected signals, and 1 indicates a very high change. But for posture detection, *Argosleep* requires discrete binary states: Rest or toss-turn. Further, due to sensitivity of mmWave signals to minute changes in the environment, 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 4, where the turn time duration estimated by mmWave signals is much larger than the Kinect output.

2) *Improving Detection with a Two-States HMM:* To overcome this challenge and improve the detection accuracy and timing estimations, we design a lightweight two-states HMM [29]. The HMM not only converts the envelope with real-valued output between 0 to 1 to a discrete output of 0 and 1,

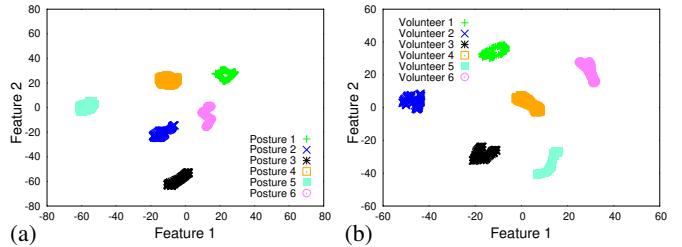


Figure 6: (a) t-SNE plot for different postures from a single volunteer. (b) t-SNE plot for different volunteers performing a same posture.

but also improves the state detection accuracy and reduces the timing errors. Figure 5(a) shows the state transition diagram of *Argosleep*'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 4 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 run-time, *Argosleep* first calculates the envelope from the reflected mmWave signals, and then uses the state transition and emission matrices and a Viterbi decoder to predict the binary states, corresponding to rest and toss-turn. Figure 5(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 either states, *Argosleep* aims to predict the sleep posture during the rest state.

D. Deep Learning based Sleep Posture Prediction

Argosleep 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 learn such a relationship based on the feature variance between distinct postures and individuals. To this end, we first analyze the behavior of reflected signals and their relationship with individuals' sleep postures.

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. To verify this intuition, we collect mmWave reflected signals from our setup with the bed from a single volunteer performing 6 different sleep postures. Then, we ask 6 volunteers (3 males and 3 females, height varying from 155 to 178 cm) to lie down on the bed in the same posture and

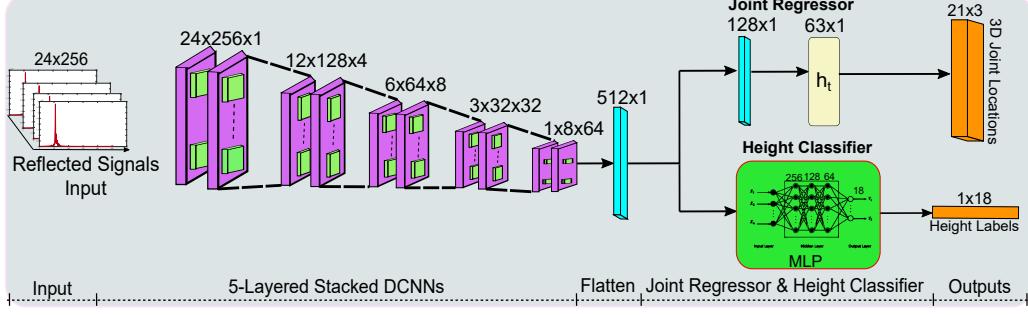


Figure 7: Argosleep’s Rest Network Architecture.

collect signals reflected from their body. 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.

Figure 6(a) and 6(b) shows the t-SNE distribution for a single volunteer performing 6 different sleep postures on the bed and for 6 volunteers performing the same posture, respectively. Clearly, we observe 6 unique feature clusters for both the cases: Such unique clusters indicate that the input signal carries enough signature so that a learning model should be able to effectively learn and distinguish them using mmWave reflected signals.

2) Argosleep’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 Deep Convolutional Neural Network (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 with her height [30]; so, Argosleep could constrain and finetune 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 [31], and 3D pose of one joint is usually constrained by its parent’s pose. So, the 3D location output of a child should be conditioned on its parent joint to ensure the distance between the parent and child is always fixed, across all postures. We now describe the network components in detail.

► **Rest Network Architecture:** Figure 7 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

Table I: Joint Regressor network parameters for Argosleep. C_j: *j*th Convolutional layer; S_i: *i*th Stack in *j*th Convolutional layer; FL: Flatten layer; FC: Fully Connected layer; Act. Fcn.: Activation Function. LReLU: Leaky ReLU.

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				
Filter Size	6x6	6x6	6x6	6x6	6x6				
Dilation	2x2, 1x1								
Act. Fcn.	LReLU	LReLU	LReLU	LReLU	LReLU	LReLU	LReLU	Linear	

the spatial features relevant to the network by observing the non-linear correlation between input-output pairs [32]. 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 finetuning processes that 5-layers of stacked convolution with 2 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 [33]. We also apply 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 I 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 [34], incorporating height information can make the model generalizable to many individuals with very little or no finetuning. A user could input her ground truth height to the monitoring system, and Argosleep 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 two fold. *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 [35]), 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. MLP in the Height Classifier comprises 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 to ensure that the network converges to an optimal value. 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 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 to the entire network, and we will discuss their choice in Section IV.

IV. IMPLEMENTATION

Hardware Platform: We implement and evaluate *Argosleep* 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 [21], 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 [22], that collects the ground truth depth images and 3D location of body joints at a frame rate of 30 fps. We follow Section III-B to preprocess the datasets. Figures 8(a–b) show our experimental setup similar to at-home bedroom setting with a Queen-sized bed placed at 2.5 m from the monitoring equipment. The setup has two antenna arrays arranged in 3×4 and 4×3 configurations that resolves reflection points in azimuth, depth, and elevation,

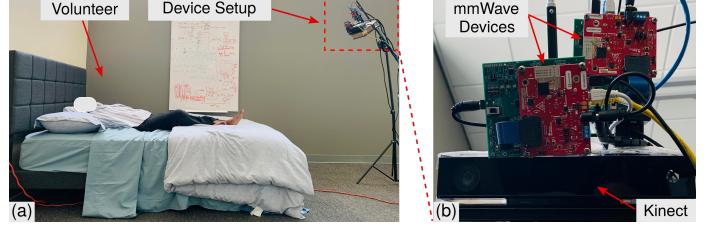


Figure 8: (a) Experimental setup with the devices and the bed, where a volunteer is performing a sleep posture. (b) Hardware platform with two 77–81 GHz mmWave devices and an RGB-D camera.



Figure 9: 5 broad categories of sleep posture in *Argosleep*.

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. We implement *Argosleep* offline on Matlab and Python environments running on host PC and GPU servers.

Real Data Collection: We collect datasets from 8 volunteers (age: 19–30 years, M/F percentage: 62.5/37.5, height range: 155–178 cm) performing several variations of the 5 sleep postures on the bed (Figure 9). The background in our experimental 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 that Kinect starts detecting the joints. Then, the volunteer is required to lie down in a posture till the end of the experiment. We only consider those frames where the body joints are successfully detected by the Kinect for our datasets.

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 finetune the Joint Regressor using the loss from this classifier (Section III-D2). For heights beyond this range, little finetuning can help the network to better understand the relation. In total, we have nearly 40 K samples with a raw data size of 16.7 GB of toss-turn events, and 27,382 static sleep postures with a processed data size of 2.5 GB with 5 postures from 8 volunteers of diverse ages, gender, and height.

Network Training: We train *Argosleep*'s Rest Network by exploring different parameter settings to ensure convergence to a near-optimal value. *Argosleep*'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 2500, then monitor the training process till the total loss function shows no improvement for consecutive 30 epochs. *Then*, we explore different optimizers, such as, *Adam*, *Rmsprop*, *SGD*, etc., 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. *Argosleep*'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. The Rest network is implemented in Python with TensorFlow 2.4 using Spyder IDE and Anaconda version 4.10.3 distribution. Our training time varies between 6 to 10 hours for completion in a GPU server with 2 NVIDIA RTX A6000 nodes.

V. Argosleep EVALUATION

Evaluation Summary: (1) *Argosleep*'s HMM-Viterbi detects toss-turns with a median accuracy $\sim 85\%$ and median precision, recall, and F1-score of 0.97, 0.88, and 0.92, respectively, indicating high accuracy and low false positives. *Argosleep* 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 median. (2) *Argosleep*'s Rest Network predicts the 3D location of body joints of a person's 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.

A. Toss-Turn Detection Results

► **State Detection Performance:** We first evaluate the effectiveness of *Argosleep*'s toss-turn detection modules. Since it is hard to control the number of toss-turns during the actual sleep, and obtain a reasonably-sized dataset, we obtain 19,386 state observations from 30 datasets collected from a volunteer who mimics the toss-turn events with posture changes within a short period of 20 to 30 seconds to generate one dataset. We generate the input-output pairs of mmWave reflected signals and 3D location of body joints and also, corresponding Kinect depth images to identify the ground truth rest or toss-turn states. We find the ground truth toss-turn by applying a fixed mask and calculating the pixel-to-pixel difference in successive depth images and then, finding the energy in residual depth. *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.

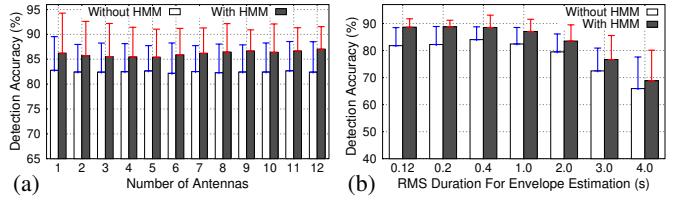


Figure 10: Rest and toss-turn detection accuracy: Effects of (a) the number of antennas. (b) different RMS duration for envelope estimation. The bar and errorbar represent the median and standard deviation across 19,386 states.

Figure 10(a) shows the state detection accuracy *w.r.t.* the Kinect-based ground truth, across different number of antennas. We use the observation from multiple antennas and take median votes to decide between the output binary states. Clearly, *Argosleep*'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 since a single antenna, with a large beamwidth, can cover the whole bed area. We also 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 10(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, *Argosleep* performs consistently better with HMM-Viterbi, and RMS duration of 1 s shows 88% detection accuracy on the median. Figure 11(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 *Argosleep* is not only accurate but also has low false detection rates.

► **Toss-Turn Timing Parameters:** Next, we evaluate *Argosleep*'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 *Argosleep* 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 *Argosleep* 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 from the ground truth and *Argosleep*.

Figure 11(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 *Argosleep* did not miss any detection. *Second*, without HMM-Viterbi,

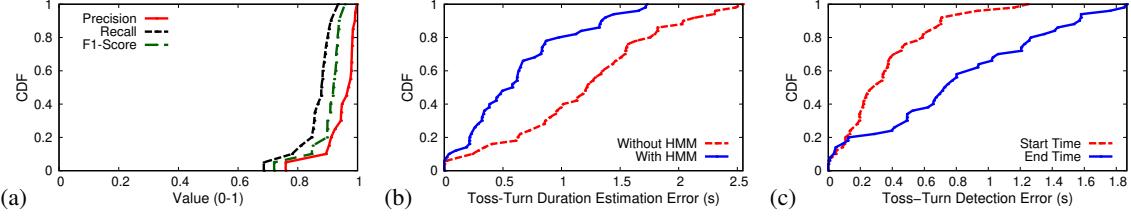


Figure 11: (a) Precision, recall, and F1-score for the state detection with HMM. (b) Distribution of errors in toss-turn duration estimation, with and without HMM. (c) Distribution of errors in toss-turn start and end times detection with HMM.

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, *Argosleep* always predicts the duration within 1.7 s of the ground truth across all our observed events. Figure 11(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 *Argosleep* is accurate within 1 s and 1.5 s to detect the start and end of the event, respectively.

B. Sleep Posture Prediction Results

► **Error in Estimating 3D Location of Body Joints:** We now evaluate the performance of *Argosleep*'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 ~9730 samples collected from three volunteers, (2 females and 1 male, height varying from 155 cm to 178 cm) with the lowest and highest height among all volunteers, performing 5 different sleep postures and their variations, and then evaluate the performance across all 8 volunteers. After synchronizing and resampling, we label the volunteers' height into discrete categories following Section IV. Then, we randomly select ~8700 samples for training and ~1030 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 12(a) shows the performance of *Argosleep*, 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). To investigate this issue further, we plot the aggregated errors from all joints and separate them in terms of the sleep postures. Figure 12(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. But the joints that are critical to facilitate a sleep posture monitoring application can be predicted accurately by *Argosleep*. Figure 13 shows top-view of skeletons for various sleep postures predicted by *Argosleep*. These results demonstrate that *Argosleep* can predict the 3D location of body joints accurately.

► **Effect of Height Classifier:** *Argosleep* uses the output from the Height Classifier to finetune its Joint Regressor to improve its generalization ability and 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 ~6500 samples collected from three volunteers, and test it on another set of ~2000 samples. Furthermore, we build the Height Classifier following Section III-D2 into the network and feed its loss function to finetune the output of the Joint Regressor. Then, we evaluate the performance with the same set of training and testing samples.

Figure 12(c) shows the performance of the Rest Network with and without the Height Classifier. We observe that *Argosleep* 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, enable better joint location estimation from the reflections.

► **Effect of Number of Volunteers:** To evaluate the generalizability of *Argosleep*'s Rest Network for diverse volunteers, we now perform an ablation study. Here, we would like to understand the performance and amount of finetuning required for new, unseen volunteers for *Argosleep*. To this end, we randomly select 2000 test samples from 8 volunteers, including all 5 sleep postures, with 250 samples from each volunteer. These are unseen data for *Argosleep*'s Rest Network. We then create a training set of ~3000 samples from one volunteer and train *Argosleep*'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 2 new volunteers' datasets and finetune the base model, and test on the same set of test samples for 8 volunteers.

Figure 14 shows the performance of *Argosleep* with different levels of finetuning. 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

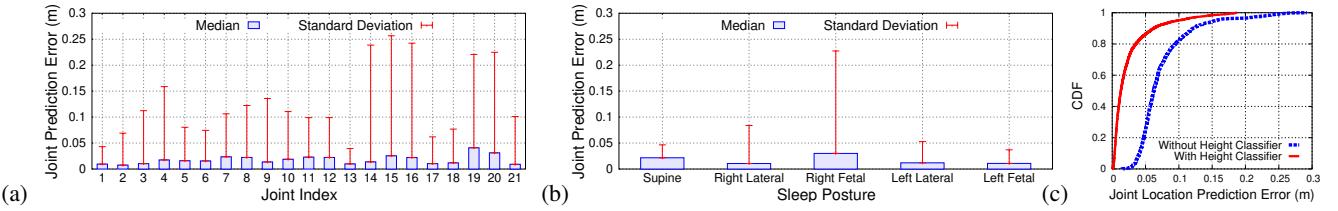


Figure 12: (a) Argosleep’s prediction errors for 21 joints across 5 poses for 3 volunteers. (b) Argosleep’s prediction errors across all joints, summarized for individual sleep postures. (c) Height Classifier improves the performance of Rest Network.

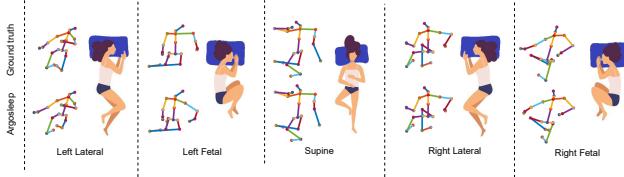


Figure 13: Argosleep predicts the location of key body joints, similar to vision-based systems, using mmWave signals.

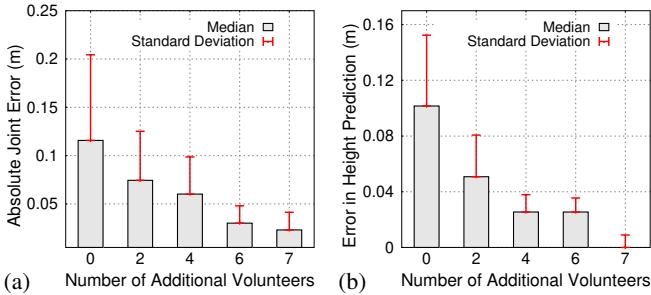


Figure 14: Argosleep generalizes across multiple volunteers with only a little finetuning. (a) Joint and (b) Height errors.

is highly 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 finetuning the network with 2 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 Argosleep can learn feature associations between individuals and their sleep posture 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 finetuning.

In summary, Argosleep identifies the toss-turn events accurately and predicts 3D location of body joints with accuracy on par with the existing vision-based system.

VI. RELATED WORK

RF-based Sleep Posture Detection: Significant research efforts have been directed to understanding and inferring sleep posture using RF signals. [16] uses signals from Wi-Fi like device to identify the angular orientation of a person to determine sleep posture. [17] also uses Wi-Fi to exploit the fine-grained channel information to capture the minute movements caused by breathing and heartbeats. [18], [36] 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 [37]–[39]. In particular, extracting skeletal information has been the main focus as it provides visual information about the 3D pose of a person [40]–[43]. 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 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 states, and changes in postures take place abruptly. Argosleep 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.

VII. CONCLUSION

We present Argosleep, a single person sleep posture monitoring system that detects the toss-turn events, and predicts the 3D location of body joints during sleep. Argosleep 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 Argosleep generalizes to multiple volunteers with little finetuning and works well for different sleep postures. We plan to extend Argosleep to monitor sleep postures for two persons, and collect datasets for long durations to evaluate its end-to-end performance. We believe Argosleep can unlock the potential of 5G mmWave systems, such as home wireless routers, in enabling privacy-noninvasive, high-quality at-home sleep monitoring systems.

ACKNOWLEDGMENTS

We sincerely thank the reviewers for their comments and feedback. This work is partially supported by the NSF under grants CNS-1910853, CAREER-2144505, and MRI-2018966.

REFERENCES

- [1] Lisa Matricciani and Yu Sun Bin and Tea Lallukka and Erkki Kronholm and Melissa Wake and Catherine Paquet and Dorothea Dumuid and Tim Olds , “Rethinking the Sleep-Health Link,” *Sleep Health*, vol. 4, no. 4, 2018.
- [2] Jason J. Liu and Wenyao Xu and Ming-Chun Huang and Nabil Alshurafa and Majid Sarrafzadeh and Nitin Raut and Behrooz Yadegar, “Sleep Posture Analysis Using a Dense Pressure Sensitive Bedsheet,” *Pervasive and Mobile Computing*, vol. 10, 2014.
- [3] The SUN, “DREAM TEAM People Are Arguing About What is The Best Position to Sleep in but Which One Are You?” 2020. [Online]. Available: <https://www.thesun.co.uk/fabulous/10791045/sleep-position-best-people-argue-night/>
- [4] Lee, Hedok and Xie, Lulu and Yu, Mei and Kang, Hongyi and Feng, Tian and Deane, Rashid and Logan, Jean and Nedergaard, Maiken and Benveniste, Helene, “The Effect of Body Posture on Brain Glymphatic Transport,” *Journal of Neuroscience*, vol. 35, no. 31, 2015.
- [5] Tamis Pin and Beverley Eldridge and Mary P Galea, “A Review of the Effects of Sleep Position, Play Position, and Equipment Use on Motor Development in Infants,” *Developmental Medicine and Child Neurology*, vol. 49, no. 11, 2007.
- [6] Gustavo Desouzart and Rui Matos and Filipe Melo and Ernesto Filgueiras, “Effects of Sleeping Position on Back Pain in Physically Active Seniors: A Controlled Pilot Study,” *Work*, vol. 53, no. 2, 2013.
- [7] Simon A. Joosten and Denise M. O’Driscoll and Philip J. Berger and Garun S. Hamilton, “Supine Position Related Obstructive Sleep Apnea in Adults: Pathogenesis and Treatment,” *Sleep Medicine Reviews*, vol. 18, no. 1, 2014.
- [8] Stefani, Ambra, and Birgit Hgl, “Sleep in Parkinson’s disease,” *Neuropsychopharmacology*, vol. 45, no. 1, 2020.
- [9] Sleep Foundation, “Best Sleeping Positions,” 2022. [Online]. Available: <https://www.sleepfoundation.org/sleeping-positions>
- [10] Johns Hopkins, “Choosing the Best Sleep Position,” 2022. [Online]. Available: <https://www.hopkinsmedicine.org/health/wellness-and-prevention/choosing-the-best-sleep-position>
- [11] MFI Medical, “Embla N7000 PSG System - Certified Refurbished,” 2022. [Online]. Available: <https://mfimedical.com/products/embla-n7000-psg-system>
- [12] Shinjae Kwon and Hojoong Kim and Woon-Hong Yeo, “Recent Advances in Wearable Sensors and Portable Electronics for Sleep Monitoring,” *iScience*, vol. 24, no. 5, 2021.
- [13] Amazon, “TACTILUS Mattress Pressure Mapping Sensor System Tactile Force Bed Software Body,” 2022. [Online]. Available: <https://www.amazon.com/TACTILUS-Mattress-Pressure-Mapping-Software/>
- [14] Lee Jaehoon and Min Hong and Sungyong Ryu, “Sleep Monitoring System Using Kinect Sensor,” *International Journal of Distributed Sensor Networks*, 2015.
- [15] Liu, Shuangjun and Ostadabbas, Sarah, “A Vision-Based System for In-Bed Posture Tracking,” in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017.
- [16] Yue, Shichao and Yang, Yuzhe and Wang, Hao and Rahul, Hariharan and Katabi, Dina, “BodyCompass: Monitoring Sleep Posture with Wireless Signals,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 2, 2020.
- [17] Liu, Jian and Chen, Yingying and Wang, Yan and Chen, Xu and Cheng, Jerry and Yang, Jie, “Monitoring Vital Signs and Postures During Sleep Using WiFi Signals,” *IEEE Internet of Things Journal*, vol. 5, no. 3, 2018.
- [18] Zhang, Feng and Wu, Chenshu and Wang, Beibei and Wu, Min and Bugos, Daniel and Zhang, Hangfang and Liu, K. J. Ray, “SMARS: Sleep Monitoring via Ambient Radio Signals,” *IEEE Transactions on Mobile Computing*, vol. 20, no. 1, 2021.
- [19] Moein Enayati and Marjorie Skubic and James M Keller and Mihail Popescu and Nasibeht Zanjirani Farahani, “Sleep Posture Classification Using Bed Sensor Data and Neural Networks,” in *Annu Int Conf IEEE Eng Med Biol Soc.*, 2018.
- [20] Zhang, Feng and Wu, Chenshu and Wang, Beibei and Liu, K. J. Ray, “mmEye: Super-Resolution Millimeter Wave Imaging,” *IEEE Internet of Things Journal*, vol. 8, no. 8, 2021.
- [21] Texas Instruments, “IWR1443 Single-Chip 76-GHz to 81-GHz MmWave Sensor Evaluation Module,” 2020. [Online]. Available: <https://www.ti.com/tool/IWR1443BOOST>
- [22] Gamesto, “Kinect,” 2022. [Online]. Available: <https://www.gamestop.com/gaming-accessories/controllers/xbox-one/products/microsoft-kinect-for-xbox-one/10115985.html>
- [23] Meta, Adriano and Hoogboom, Peter and Lighart, Leo P., “Signal Processing for FMCW SAR,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, 2007.
- [24] NETGEAR, Inc., “Nighthawk X10 Smart WiFi Router,” 2022. [Online]. Available: <https://www.netgear.com/landings/ad7200/>
- [25] Airfide Networks, “Airfide Brings High Performance Home and Enterprise 5G-NR Wireless,” 2022. [Online]. Available: <https://airfidenet.com/>
- [26] Kazemi, Mahmoud and Kavehvash, Zahra and Shabany, Mahdi, “K-Space Analysis of Aliasing in Millimeter-Wave Imaging Systems,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 3, 2021.
- [27] Wang, Xueyi and Ellul, Joshua and Azzopardi, George, “Elderly Fall Detection Systems: A Literature Survey,” *Frontiers in Robotics and AI*, vol. 7, 2020.
- [28] Yisak Kim and Juyoung Park and Hyungsuk Kim, “Signal-Processing Framework for Ultrasound Compressed Sensing Data: Envelope Detection and Spectral Analysis,” in *MDPI Applied Sciences*, 2020.
- [29] Daniel Jurafsky and James H. Martin, *Speech and Language Processing*, 3rd ed. Pearson Prentice Hall, 2021.
- [30] Chen, Tianlang and Fang, Chen and Shen, Xiaohui and Zhu, Yiheng and Chen, Zhili and Luo, Jiebo, “Anatomy-aware 3D Human Pose Estimation with Bone-based Pose Decomposition,” 2020. [Online]. Available: <https://arxiv.org/abs/2002.10322>
- [31] Blazevic, Martin and Brkic, Karla and Hrkac, Tomislav, “Towards Reversible De-Identification in Video Sequences Using 3D Avatars and Steganography,” 2015. [Online]. Available: <https://arxiv.org/abs/1510.04861>
- [32] Albawi, Saad and Mohammed, Tareq Abed and Al-Zawi, Saad, “Understanding of a convolutional neural network,” in *2017 International Conference on Engineering and Technology (ICET)*, 2017.
- [33] Karen Simonyan and Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” 2015. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [34] Porter, A., “The Prediction of Physique from the Skeleton.” *International Journal of Osteoarchaeol*, vol. 9, 1999.
- [35] World Population Review, “Average Height by State 2022,” 2022. [Online]. Available: <https://worldpopulationreview.com/state-rankings/average-height-by-state>
- [36] Liu, Xuefeng and Cao, Jannong and Tang, Shaojie and Wen, Jiaqi, “Wi-Sleep: Contactless Sleep Monitoring via WiFi Signals,” in *2014 IEEE Real-Time Systems Symposium*, 2014.
- [37] Lee, Gawon and Kim, Jihie, “Improving Human Activity Recognition for Sparse Radar Point Clouds: A Graph Neural Network Model with Pre-Trained 3D Human-Joint Coordinates,” *Applied Sciences*, vol. 12, no. 4, 2022.
- [38] Yu, Chengxi and Xu, Zhezhuang and Yan, Kun and Chien, Ying-Ren and Fang, Shih-Hau and Wu, Hsiao-Chun, “Noninvasive Human Activity Recognition Using Millimeter-Wave Radar,” *IEEE Systems Journal*, 2022.
- [39] Xue, Hongfei and Ju, Yan and Miao, Chenglin and Wang, Yijiang and Wang, Shiyang and Zhang, Aidong and Su, Lu, “mmMesh: Towards 3D Real-Time Dynamic Human Mesh Construction Using Millimeter-Wave,” in *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, 2021.
- [40] Sengupta, Arindam and Jin, Feng and Zhang, Renyuan and Cao, Siyang, “mm-Pose: Real-Time Human Skeletal Posture Estimation Using mmWave Radars and CNNs,” *IEEE Sensors Journal*, vol. 20, no. 17, 2020.
- [41] Arindam Sengupta and Siyang Cao, “mmPose-NLP: A Natural Language Processing Approach to Precise Skeletal Pose Estimation using mmWave Radars,” 2021. [Online]. Available: <https://arxiv.org/abs/2107.10327>
- [42] Wang, Qing and Wang, Kai and Chen, Wai, “CLGNet: A New Network for Human Pose Estimation Using Commodity Millimeter Wave Radar,” in *2020 3rd International Conference on Algorithms, Computing and Artificial Intelligence*, 2020.
- [43] Ding, Wen and Cao, Zhongping and Zhang, Jianxiong and Chen, Rihui and Guo, Xuemei and Wang, Guoli, “Radar-Based 3D Human Skeleton Estimation by Kinematic Constrained Learning,” *IEEE Sensors Journal*, vol. 21, no. 20, 2021.