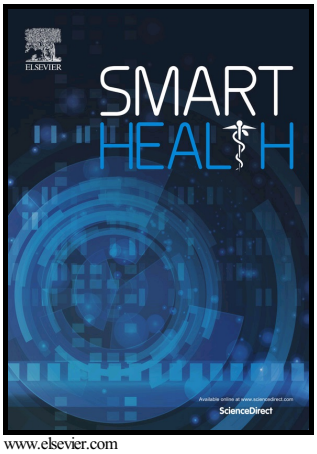

Volumetric Reconstruction of Thermal-depth Fused 3D Models for Occluded Body Posture Estimation

Shane Transue, Phuc Nguyen, Tam Vu, Min-Hyung Choi

PII: S2352-6483(17)30103-4
DOI: https://doi.org/10.1016/j.smhl.2018.03.003
Reference: SMHL27

To appear in: *Smart Health*

Received date: 13 October 2017
Revised date: 15 January 2018
Accepted date: 3 March 2018

Cite this article as: Shane Transue, Phuc Nguyen, Tam Vu and Min-Hyung Choi, Volumetric Reconstruction of Thermal-depth Fused 3D Models for Occluded Body Posture Estimation, *Smart Health,* https://doi.org/10.1016/j.smhl.2018.03.003

Volumetric Reconstruction of Thermal-depth Fused 3D Models for Occluded Body Posture Estimation

Shane Transue[†], Phuc Nguyen[‡], Tam Vu[‡], and Min-Hyung Choi[†]

[†]*University of Colorado Denver,* [‡]*University of Colorado Boulder*

Abstract

Reliable and effective occluded skeletal posture estimation is a challenging problem for vision-based modalities that do not provide inter-surface imaging. These include both visible-light and depth-based devices that all existing pose estimation techniques are derived from. This fundamental limitation in skeletal tracking is due to the inability of these techniques to penetrate occluding surface materials to derive joint configurations that are not directly visible to the camera. In this work, we present a new method of estimating skeletal posture in occluded applications using both depth and thermal imaging through volumetric modeling and introduce a new occluded ground-truth tracking method inspired by modern motion capture solutions for tracking occluded joint positions. Using this integrated volumetric model, we utilize Convolutional Neural Networks to characterize and identify volumetric thermal distributions that match trained skeletal posture estimates which includes disconnected skeletal definitions and allows correct posture estimation in highly ambiguous cases. We demonstrate this approach by accurately identifying common sleep postures that present challenging cases for current skeletal joint estimation techniques and evaluate the use of volumetric thermal models for various sleep-study related applications.

1. Introduction

Accurate and reliable occluded skeletal posture estimation presents a fundamental challenge for vision-based methods that rely on depth-imaging [1] to form accurate skeletal joint estimations [2, 3] extracted from carefully selected depth features [4]. Modern skeletal estimation techniques provide a solid foundation for skeletal estimations of users in non-confined areas with no visual occlusions, however these techniques are not well suited for applications that include visual obstructions such as respiration and sleep posture studies where patients are heavily occluded by both clothing and common forms of bedding. While recent depth-based imaging methods [5, 6] have begun exploring how to solve this problem using existing depth-based methodologies, they still lack two primary fundamental components of occluded posture estimation: (1) the ability to provide an accurate ground-truth with an occluding medium present and (2) the ability to deal with extensive depth-surface ambiguities that may drastically interfere with joint position estimations required for accurate skeletal tracking. These ambiguities and direct occlusions incurred through depth imaging dictate that an individual depth surface provided by these techniques is insufficient to provide a reliable means of estimating an occluded skeletal posture, and in most cases fail to identify obscured skeletal joints. This is due to these techniques heavily relying on training-based approaches to infer joint positions and have no direct method for tracking occluded joints. In this work, we explore how the limitations of depth-imaging can be supplemented through integrating thermal imaging to introduce a new method in occluded skeletal posture estimation that includes tracking thermal signatures through an occlusion medium using volumetric modeling and how to limit the feasible spatial domain of occluded joints within this volume for more reliable *partial posture estimations* that contain completely obscured joints that can be inferred using existing techniques. We then extensively evaluate the reliability, accuracy, and feasibility of using a thermal-depth fusion technique for various sleep related applications, including the clinical significance of introducing a reliable occluded posture estimation algorithm for use with existing vision-based and radar-based [7, 8] tidal-volume estimation techniques used within numerous types of sleep studies. These respiration analysis and tidal-volume monitoring techniques require accurate skeletal posture for radar alignment [8] and joint position estimations for 4D chest respiratory model reconstructions [9].

In our novel approach to occluded skeletal posture estimation, we build off of our initial contributions within [10] to refine and evaluate our three primary contributions: (1) we present a thermal-based marker system for obtaining an occluded skeletal posture estimate derived from modern motion capture techniques for defining a ground-truth for occluded skeletal joint positions, (2) develop a volumetric representation of patient's thermal distribution within an occluded region to provide a highly detailed posture volume reconstruction and vastly reduce the search space of potential positions for occluded joints, and (3) introduce a coarse-grained skeletal posture estimation technique for identifying visually obscured joint positions. By addressing the challenges in occluded thermal imaging and introducing a robust volumetric model for posture estimation, we evaluate the proposed method by assessing its ability to correctly identify several common sleep postures and generate accurate skeletal joint positions based on an occluded patient. We then extend our evaluation to analyze the feasibility of using the system within a clinical setting and thoroughly investigate potential directions for improving upon our initial approach and implemented prototype.

Acknowledgments: This work is partially supported by the Department of Education GAANN Fellowship: P200A150283 and NSF Grant: 1602428.

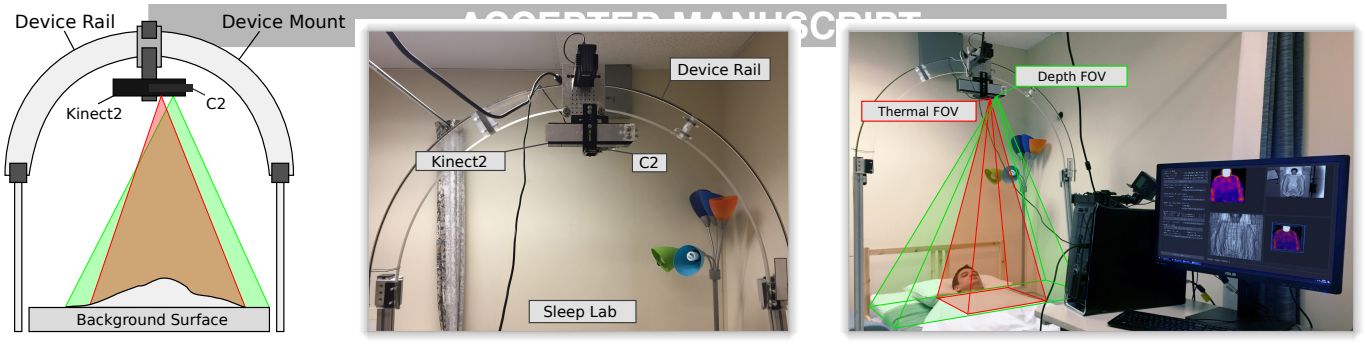

Figure 1: Experimental setup for detecting occluded skeletal joints that define a patient's posture with occlusions within a standard sleep-study. The prototype design (left) incorporates two devices: (1) a time-of-flight depth camera (Microsoft Kinect2) and (2) a thermal camera (FLIR C2), both mounted to a fixed device rail (center). This results in the fusion between these imaging modalities due to the overlapping Field-Of-View (FOV) from each device. The image streams are then fused into a thermal point-cloud within our software solution (right).

Modern digital imaging devices contain several alternative forms of imaging that utilize different wavelengths of the electromagnetic spectrum that are capable of providing information about internal skeletal structures through occlusions. However, for both practical applications and in medical practice these imaging techniques are not well suited, convenient, or safe for extended exposures over long periods of time, as is common in most sleep studies. To strike a balance between safe and reliable imaging techniques that allow us to gain information about the occluded skeletal posture of the patient, the developed hardware prototype provides a real-time posture estimation derived from both depth and thermal imaging devices that are mounted on a common device rail to provide an overlap of the Field-Of-View (FOV) that mimics a stereoscopic fusion of the two imaging modalities. From this hardware setup, we can utilize existing methods within image stereoscopy [11] to *fuse* the two image streams into a coherent thermal point-cloud. From these parallel data streams, we provide a software interface that illustrates four components: (1) the infrared stream, (2) the depth image stream visualized as a 3D point-cloud, (3) the thermal image, and (4) the fusion between the 3D point-cloud and the thermal images which results in a thermal point-cloud that we record over time. This system provides the foundation we use to perform both the volumetric reconstruction of the patients posture and the main system for extracting the skeletal ground-truth measurements required for our training procedure. In Figure 1, we provide a complete overview of our experimental Sleep Lab setup used for both training and monitoring occluded skeletal postures.

2. Related Work

Skeletal posture estimation from imaging devices is a field within computer vision that has received an extensive amount of attention for several years since the introduction of widely-available depth-imaging devices. Through the development of several devices that support high-resolution depth imaging, depth-based skeletal estimation has become a robust and mature method of providing joint and bone-based skeletal estimations. Notable contributions to this work include both generations of the Microsoft Kinect, associated depth-based skeletal tracking algorithms, and the extensive set of work aimed at improving these skeletal estimations. While these existing techniques are well explored and reliable for most applications, they are inherently ineffective for posture estimations that include visual occlusions like those encountered in sleep-based studies. To assess existing methods and their inherent limitations within this application domain, we look at the most recent forms of depth-based skeletal estimation and look for related methods that attempt to tackle the problem of occluded posture estimation using these techniques.

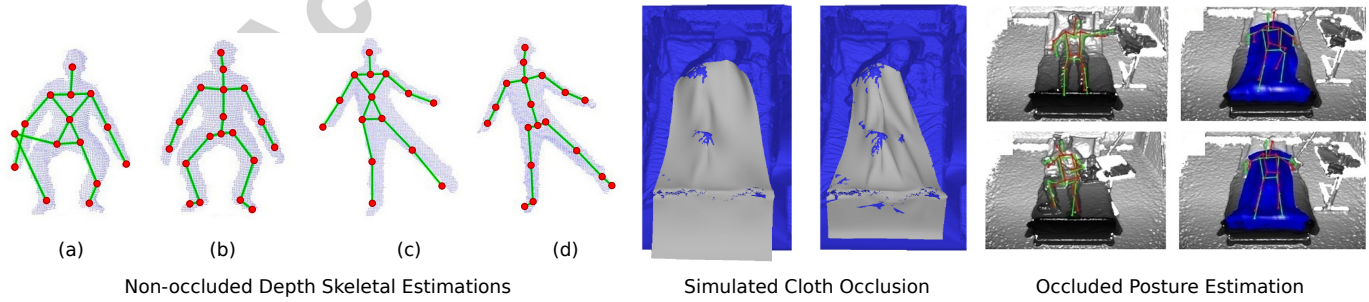

Figure 2: Skeletal posture estimations techniques using the Microsoft Kinect (a, c), and improvements (b, d) by [2]. These methods have been developed into systems that identify skeletal postures under occluding surfaces (blankets) using physical simulation (center) to infer depth feature signature patterns generated from occluding surfaces [5] to estimate skeletal postures (right).

Depth-based Skeletal Estimation. The pioneer work for depth-based skeletal estimation from a single depth image for the Microsoft Kinect devices [3, 2] utilized a combination of both depth-image body-segment feature recognition and training through Random Decision Forests (RDFs) to rapidly identify depth pixel information and their contribution to known skeletal joints and hand gestures, as introduced with the Kinect2. Modern skeletal estimation techniques are built around a similar premise and

utilize an extensive number of newer devices that provide high-resolution depth images. These techniques utilize temporal correspondence, feature extraction, and extensive training sets to quickly and robustly identify key regions within a human figure that correlate to a fixed number of joint positions that form a skeletal structure of the user. The images in Figure 2 provide an illustration of the most common skeletal configurations and associated estimation results from recent techniques. These techniques have become increasingly robust and now provide highly accurate joint estimations within the well established constraints of these approaches. These constraints minimize assumptions about the free movement of the human skeleton and provide reasonable joint movements. However, these techniques also provide a set of assumptions including: background data can be quickly segmented (removed), the user is relatively isolated within the depth image, and most importantly - the line of sight between the device and the user is not obstructed. These assumptions are integrated into the foundation of these approaches, therefore the use of these methods within sleep-based studies with occluding materials covering the patient are not valid under these constraints.

Occluded Skeletal Posture Estimation. Recent vision-based techniques have introduced an alternative method that relies on a surface prior to allow skeletal posture estimations that are recorded before the occluding medium is introduced [5]. This surface prior (depth-image) is then used as a collision model within a physical simulation of a cloth that represents the occluding surface to provide an approximation of what the underlying posture would look like given the simulated cloth model occluding the patient. However, there are several potential problems with this approach: (1) the simulated cloth under gravity model may not provide realistic behaviors such as folding and tucking, (2) body movement may modify the blanket for instances not covered in the simulation, and (3) the patient may move and create additional wrinkles, folds, layering, self-collisions, and complex interactions between the patient and the cloth model. While this method provides a good alternative for depth-imaging approaches, it is difficult to ensure that the simulated cloth is consistent with real-world deformation patterns and cannot emulate complex patent to blanket interactions that may be observed.

Alternative methods derived from signal and image processing [6] have also been introduced in an attempt to identify a patient's posture based on the spatial domain patterns that can be extracted by processing cross-sections of the bed surface using the Fast Fourier Transform (FFT). The objective of this approach is to identify the spatial patterns common to most postures and then identify them based on these traits. However, similar to other depth imaging approaches, the surface data provided through a surface point-cloud does not contain accurate information about the posture of the patient within the occluded volume. Therefore in occluded applications, the high level of surface ambiguities makes depth-based techniques ill-suited for accurately estimating skeletal joint positions, even for this form of signal processing.

Thermal Image Posture Estimation. The use of thermal imaging for skeletal posture estimation has not been extensively utilized due to the fact that thermal images do not provide a good estimate of the spatial coordinates required for skeletal joints, with the exception of 2D movement tracking. Early work presented in [12] developed a simple algorithm for detecting the skeletal structure within a two-dimensional image, but the applications of this method are limited to the 2D domain and cannot be utilized to form a 3D spatial representation of a patient's posture. Recently, there has been limited exploration into thermal-based skeletal estimation, however the technique has been used for detecting [13] and tracking generalized human behaviors [14, 15] which include movement and very generic postures such as walking, lying, and sitting. However, none of these techniques have explored combining depth and thermal imaging to improve skeletal estimates especially in cases where occlusion makes depth-only methods invalid. Additionally, many approaches, including ours, use low-cost commercially available thermal cameras, making the resolution of the images limited. In these existing techniques, this reduces the image quality and greatly increases the difficulty of extracting accurate joint estimates simply due to the hardware limitations. Therefore, to address the introduction of an occluding material within skeletal estimates, we fuse both thermal and depth imaging to provide a means of generating a thermal model of the patient's volume enclosed by the occluding medium by leveraging high-resolution depth images to provide spatial and visual fidelity to lower resolution thermal images.

3. Occluded Posture Thermal Challenges

The extensive depth of research used to provide reliable techniques for accurate joint estimates using single depth images has generated a significant number of solutions for posture estimation in occlusion-free applications for skeletal movement tracking with high degrees of freedom. With the introduction of occlusion mediums, the addition of thermal imaging to assist in the identification of a patient's skeletal posture provides an intuitive extension of these techniques. However, with the introduction of markerless skeletal posture estimation and visual occlusions, thermal imaging retains an extensive set of challenges due to heat propagation, contact regions, and the potential occlusion materials that can be used to block a patient's skeletal posture. In this section we enumerate several primary challenges associated with thermal imaging that greatly complicate thermal-based skeletal estimation and explore how the fusion between this modality and depth-imaging can alleviate these potential problems.

Occluded Ground-truth Estimation. One of the prominent challenges with establishing an algorithm for occluded posture estimation stems from the inability of current vision-based approaches to define an accurate ground-truth of an occluded skeletal posture. This is due to the use of imaging wavelengths that are blocked by specific wavelength opaque surfaces which makes most vision-based techniques inadequate for visualizing internal structures occluded by surface materials. This includes both the visible spectrum of color images and the short infrared wavelengths used for depth imaging. Therefore, for skeletal posture estimation with surface occlusions, the process of determining a ground-truth estimation of the patient's posture is in most

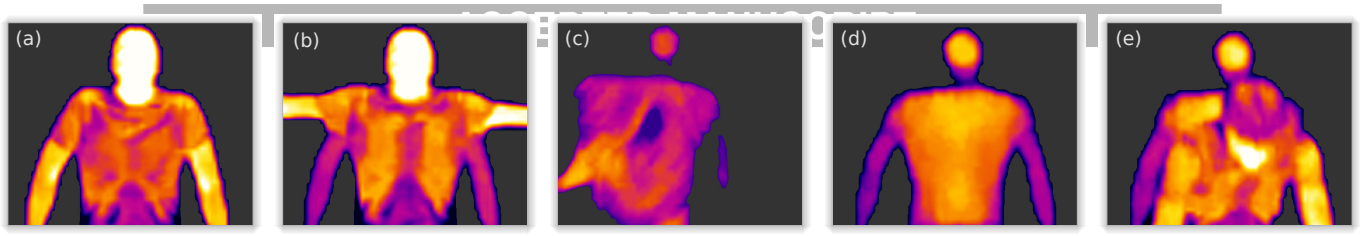

Figure 3: Skeletal posture estimation challenges associated with thermal imaging. The image in (a) illustrates an ideal non-occluded thermal image but illustrates non-uniform thermal distribution of a patient's thermal signature, (b) provides an illustration of heat marks left by a patient's arm movements, (c) illustrates thermal ambiguities of the patient during motion, (d) illustrates the patient's residual heat left when the patient has been removed, and (e) illustrates ambiguous movement where the posture may be interpreted as multiple skeletal components.

instances difficult or completely intangible. This eliminates the possibility of using traditional ground-truth tracking techniques, like those used within common motion-capture studios, to establish large sets of skeletal training data that are used to provide the critical link between depth image features and joint position estimates. Therefore we look towards how we can fuse thermal and depth imaging and provide a new method inspired by these traditional techniques to establish a ground-truth estimate that can be directly measured from occluded joint positions.

Contact Regions. The thermal conductivity exhibited by a material near a heat emitting source can be simplified and modeled using two different thermal transfer states: (1) a non-contact state which defines a scalar distance that separates the source and the receiving material and (2) a contact state where heat transfer is greatly increased due to the thermal contact conductance between the two materials. In the first case, thermal conductance is reduced and defined as a function of the distance between the emitting surface and the receiving material which depends on the ambient temperature, temperature of the two objects, and the material composition of both objects. This is true for the second case, however due to the contact surface, the thermal conduction is greatly increased, leading to a substantial increase in thermal intensity. Therefore to accurately describe an object's thermal contacts, the shape of its surface, and emission intensity from a thermal image, the physical properties of all materials must be precisely modeled, which is impractical for clinical applications.

Limb Occlusions. As with all single perspective depth-based posture estimations, occlusions made by specific poses incur constraints on the accuracy of the skeletal posture estimation due to limbs occluding other joints within the depth image. Within depth imaging this is handled by introducing joint states that identify when a joint position is accurately known, or if the joint position is not directly known, but can still be inferred. However, with the introduction of an occluding material *tent-effect*, in which the occluding medium creates an occlusion area larger than a traditional limb occlusion, this effect will contribute to a much more significant loss of information about other skeletal joints due to the increased occlusion volume introduced by the shadow of the occluded material within the depth image. Since this phenomena occurs in combination with large contact regions and large depth extrusions, these instances can be detected using thermal-depth fusion.

Intractable Heat-to-Surface Modeling. Identifying and generating an accurate surface model exclusively through the use of thermal imaging is an under-determined inverse physics problem. This is because there is inherently an ambiguous relationship between the measured thermal intensity and the emission surface that cannot be directly reconstructed. Thus depth-imaging remains a prominent requirement for spatial models of thermal distributions.

Non-uniform Heat Distributions. The thermal signature of the human body has a substantial natural variation across the surface of the skin that contributes to non-uniform heat distributions. The premise of any thermal-based approach to skeletal estimation assumes that the emission of thermal energy from the surface of the skin is sufficient to separate from both the background and other materials near and in contact with the skin; however due to the non-uniform distribution of heat through different skin regions and material coverage, thermal intensities are ambiguous between the patient's skin and surrounding materials.

Movement and Residual Heat. As a challenge uniquely associated with thermal imaging, thermal contact and residual heat play a critical role in the image analysis of patient postures. During the movement event and for a short period of time after the movement, thermal intensities may indicate false positives in posture estimations due to residual heat. Depending on the contact surface material properties and the duration of the contact, residual thermal signatures can generate significant misleading features that can be mistakenly identified as part of the patients posture. One prominent example of this problem is illustrated in Figure 3 (b), where the thermal image depicts four arms instead of two. While this seems ambiguous, the differentiation between the correct arm positions and their previous location (shown by their residual heat), is directly solved by analyzing depth features at these locations.

Multi-Layer Occlusions. The apparent thermal distribution of an occluded surface is directly influenced by both the distance and temperature of the emitting surface, however the number of occluding material layers between the thermal device and the emission source introduces additional erroneous ambiguities. As materials are placed on the patient, including clothes and bedding, the materials may overlap in unpredictable ways leading to sharp distinct features within the thermal image.

Occlusion Material. Material properties of the occluding surface greatly dictate the thermal distribution and resulting surface signature that can be identified by the thermal camera. Depending on the material type, thickness, and heat propagation characteristics of the material, this can play a significant role within how accurately a posture can be identified and how long the thermal signature remains as residual heat.

4. Method Overview

To provide a reliable means of estimating occluded skeletal postures in any vision-based technique, the proposed method must address the challenges presented by the data acquisition methods used create a solid foundation for performing accurate joint estimations. An immediate extension to current depth-based skeletal estimation techniques is the integration of thermal data to both identify and refine potential joint locations by analyzing thermally intense regions of the body and limiting ambiguities within the depth image to provide better joint estimates within the occluded region. However, while this approach of combining both depth and thermal image information alleviates some of the challenges and ambiguities associated with depth-imaging, it also incurs the numerous thermal challenges listed within Section 3. Therefore to provide a reliable posture estimation algorithm based on these imaging methods, we mitigate the challenges introduced by each device by forming a new thermal-volumetric model of the patient's body that can provide a robust foundation for thermal-based skeletal joint estimates.

4.1. Thermal Volumetric Posture Reconstruction

Volumetric reconstruction for posture estimation refers to the process of identifying and generating the extent and geometric characteristics of the patient's volume within the loosely defined region constrained by a depth-surface. This occluded region within the surface will be used to provide what we define as the *posture-volume* of the patient. This volume is strictly defined as the continuous region under the occluding surface that contains both the patient and empty regions surrounding the patient that are visually obscured. To define a posture estimate based on this volumetric model, we associate a fixed set of correlated skeletal joint positions within the observed thermal distribution of this volume. This allows a skeletal estimate to be identified from a known (trained) thermal distribution which represents the patient's posture under the occluding medium. Figure 4 provides an overview of this ideal posture model, the discrete volume approximation, and skeletal joint structure defined by this model.

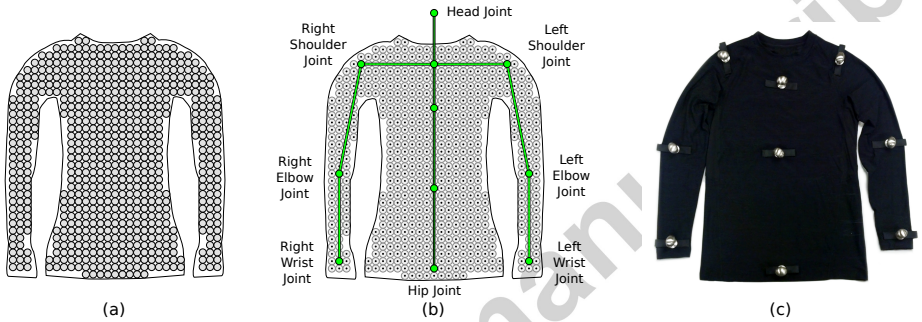

Figure 4: Volumetric reconstruction of an ideal skeletal posture. The image in (a) illustrates a discrete approximation of the patient's volume. The image in (b) provides an illustration of the mapping between a voxel representation (black dots) of this volumetric data and the ground-truth skeletal estimate of the posture (illustrated as a set of joints and associated bones).

This model shifts the foundation of the skeletal estimation from identifying isolated joints in the two-dimensional imaging domain to a three-dimensional voxel model that describes both the volume of the occluded region containing the patient and thermal distribution within this volume due to the heat radiated by the patient's skin. This form of modeling provides a complete 3D image of the patient's posture within the occluded region as an identifiable thermal distribution. We then correlate these pre-defined joints with our motion-tracking inspired ground-truth measurements that we collect during training from our thermal-joint suit shown in Figure 4 (c). This provides us with a direct correlation between ground-truth joints and their thermal distribution. This method could also be directly used for skeletal tracking, however, heated elements within an insulated volume may cause safety concerns, tangled wires or power sources are inconvenient (uncomfortable), and the non-contact complications from using this method directly limits the appeal for a non-intrusive sleep study.

The development of the volumetric posture model is motivated from three primary observations based on patient thermal images: (1) the process of identifying joint positions from thermal images projected onto the depth surface is highly unreliable due to contact region ambiguities, layering, and non-uniform heat distributions, (2) intense thermal regions within the image are generated by both joints and arbitrary locations on the patient's body, and (3) joints that have a separation distance between the patient's skin and the occluding material may be visually and thermally occluded, but reside within this volume. Due to these reoccurring conditions that are not well handled by existing methods, the proposed method is based on creating a correlation between the patient's thermal distribution and associated skeletal posture. Based on this correlation, if the known skeletal joint positions are provided for the observed distribution, we can estimate the patient's skeletal posture even when the subject is highly occluded, has several ambiguous joint positions, or when skeletal joints cannot reliably inferred.

4.2. Algorithm Overview

The premise of this approach is to reconstruct the unique volumetric thermal distribution of the patient and correlate this posture signature with an associated set of joints that defines the patient's corresponding skeletal posture. The introduction of this process provides a robust method of identifying skeletal estimates on volumetric data that contains unique thermal patterns that are more reliable than depth features within a recorded point-cloud surface. Therefore, based on our ability to reliably reconstruct

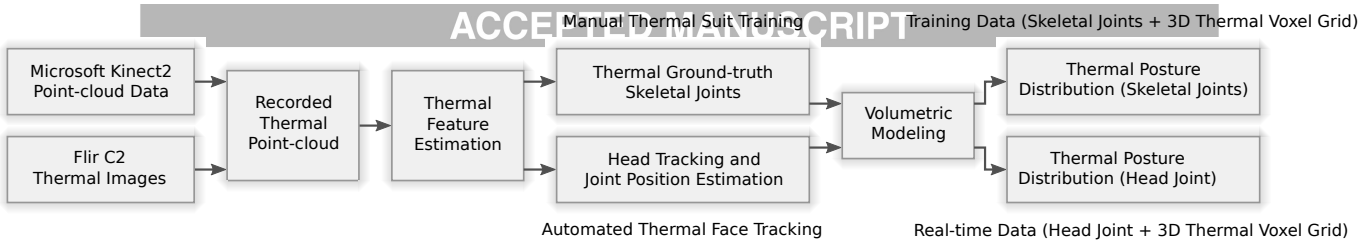

Figure 5: Overview of the proposed approach for reconstructing the volumetric thermal data that contributes to the occluded skeletal posture estimation. This includes the generation of the volumetric data with the skeletal ground-truth for training and the real-time data with the provided head joint used during the occluded posture estimation process.

this thermal distribution and associated skeletal structure, the resulting correlation is then used to populate a training model of discrete posture variants that can be used to detect a patient's subsequent postures. A high-level overview of the thermal-depth fusion process used to generate a volumetric thermal posture signature is defined in Figure 5.

The core of our technique is based on four primary components that implement the flow-process illustrated in Figure 5. This includes: (1) the generation of the thermal depth cloud through the fusion of the depth and thermal imaging devices (thermal-depth fusion), (2) the reconstruction of the patients posture using a voxel grid, (3) the propagation of the thermal values through this volume from the occluding surface, and (4) the final voxel representation of the heat distribution that is used for visualization and classification of the patient's posture. Unlike existing methods, this data is not simulated, but collected in real-time.

1. **Thermal Cloud Generation (Thermal + Depth Fusion)** Stereoscopic camera calibration has been implemented for depth and thermal imaging devices using the standard checkerboard method. This method has been extended to incorporate the addition of heat elements that correspond to the visible checkerboard pattern to identify the the intrinsics of the thermal [11] camera and their relational mapping an associated depth image. We fuse these image modalities to generate a thermal surface point-cloud that contains the occluding surface and its thermal distribution.
2. **Posture Volume Reconstruction (Sphere-packing)** The posture volume is defined as the the region occupied by the patient under the occlusion surface above the background surface (or bed surface). This occlusion surface is defined by the generated thermal-depth fusion surface and the bed surface is obtained from an initial scan (of the bed surface) without the patient present to define the back of the volume. The construction of this volume greatly reduces the skeletal joint search space and provides a high resolution approximation of the posture based on the thermal distribution.
3. **Surface Heat Propagation (Thermal Extended Gaussian Images)** From the thermal measurements of the occlusion surface, heat values are propagated into the posture volume. In this domain, the heat source is known (the patient), so it is assumed that within the volume, the thermal values increase. While this is not uniformly true, we use this heuristic to propagate heat values through the scalar field that defines the volumetric voxel grid of the posture.
4. **Volumetric Heat Distribution (Thermal Voxel Grid)** This scalar field provides a dense thermal representation of the patient's posture based on the heat propagation from the occlusion surface. This visualization is used as the basis for establishing a potential field that defines the thermal distribution of the patients posture within the volume.

This four stage process is then divided into two primary directions: (1) training for the correlation between the skeletal ground-truth and the associated thermal distribution and (2) the identification of input distributions to retrieve the patient's associated skeletal posture. To find the correlation between the ground-truth skeletal posture and its associated thermal distribution, we capture the thermal signature of our thermal-joints, label them according to their corresponding joint name, and then automatically generate the pre-defined skeletal structure. This process, repeated for various conditions, defines the *training* component of our estimation method. This procedure generates both the skeletal configuration and its associated thermal distribution. To then retrieve a skeletal estimate for a thermal volume data stream, the given thermal distribution at the current time can classified to extensively prune the search space of potential thermal distribution matches to provide the most accurate skeletal match.

5. Devices and Data Acquisition

To facilitate a practical hardware prototype that incorporates these two imaging techniques, the design incorporates two low-cost devices that provide reasonable image resolutions for sleep-based posture estimation within a controlled environment. Our prototype includes the Microsoft Kinect2 for depth imaging and the Flir C2 hand-held thermal imaging camera.

5.1. Thermal-depth Fusion Prototype

The Kinect2 provides a depth-image with a resolution of 512x424 and the C2 contains an 80x60 thermal image sensor array which is up-sampled (bicubic) to an image size of 320x240. This process does not drastically effect the quality of the thermal image due to the physical properties of the heat distribution as it propagates through the surface over time. To configure the overlapping viewable regions provided by each device, we have developed a single aluminum bracket to mount the two devices into a simple prototype as shown in Figure 6. Based on the point-cloud data provided from the Kinect2 depth-image, we integrate the thermal intensity at each point from the corresponding point within the up-sampled thermal image provided by the C2. This calibration is then used to generate the thermal-cloud of the volume enclosing the patient due to the occluding material.

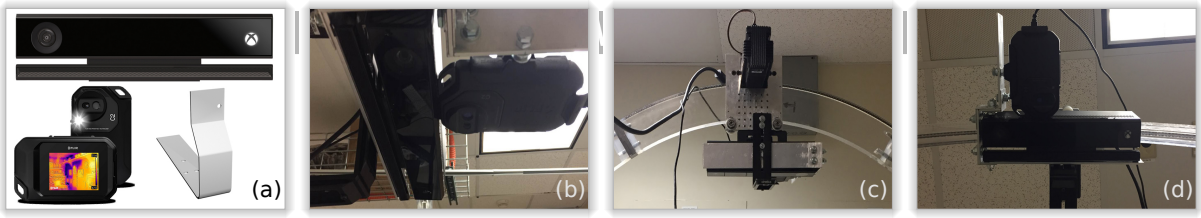

Figure 6: Thermal posture device prototype. The two devices (Kinect2, C2) are mounted with a fixed alignment provided by the bracket shown in (a). The images in (b-d) illustrate the mount attached to the bed rail with both devices.

The alignment of the images provided by these devices requires further image processing due to the vastly different field-of-view (FOV) provided by each device. Therefore we model the alignment transformation of the two camera based on a simple linear transformation as a function of the distance to the bed surface. Additionally, due to the limited FOV of the C2 device, we rotated the device by 90[deg] to provide the largest overlapping field-of-view possible. From this configuration we can sample several data streams concurrently as shown in Figure 7 to form a high-resolution thermal point-cloud. This includes the non-covered patient in (a-b) with both the infrared image in (a) and the depth image with estimated surface normals in (b). The same images are then shown for (c-d) with an occluding medium. Here, the posture is the same as the images above, however it is much more difficult to discern the exact position of the limb joints due to the uniform depth surface. Using a similar setup, we introduce thermal imaging (e-f) to illustrate the visibility of the limb joint positions even when the occluding medium is present. With the combination of this thermal data with the 3D point-cloud, the images in (g-h) provide an explicit illustration of the occluding material, the thermal distribution of the overall posture, and the penetration of the thermal signature related to the joints within the left arm, even when the occluding material is present.

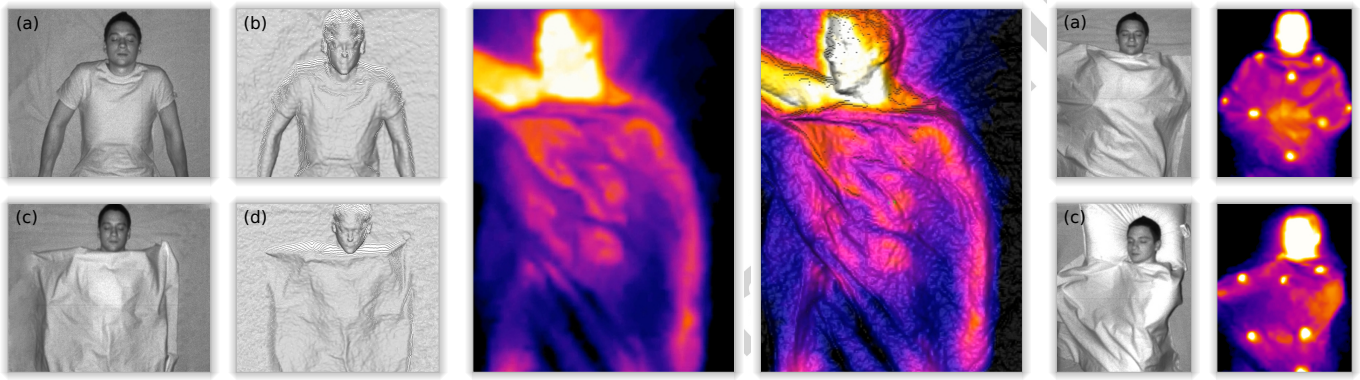

Figure 7: Thermal surface point-cloud acquisition. The sequence of images illustrate the data collected from both the Microsoft Kinect2 and Flir C2 thermal devices to obtain thermal and surface *fused* point-cloud data. The images (a-d left) illustrate the collection of the infrared and depth surfaces for both non-occluded and occluded views of the patient. The images (a-b center) illustrate thermal only and thermal-depth fusion result of our technique used to generate the volumetric model of the patient. Images (a-d right) illustrate the difficulty of identifying posture within an infrared (depth) image and the result of using thermal skeletal markers to clearly identify joint positions.

5.2. Occluded Skeletal Estimation Ground Truth

One of the prominent challenges introduced with occluded skeletal posture estimation is the inability of most vision-based techniques to provide a reliable ground-truth estimation of the patients skeletal posture while the occluding material is present. For imaging techniques, this is a direct result of the interference or complete occlusion of the patients posture due to the external surface properties of the material that are obtained through using limited regions of the electromagnetic spectrum (such as the visible or infrared wavelengths). The reflection based nature of these techniques minimizes the ability to correctly infer surface features that correctly contribute to the patient's occluded posture. While other methods utilizing these reflection-based imaging techniques have introduced interesting ground-truth workarounds for approximating the surface behavior of the occluding surface [5], this remains a significant challenge in occluded posture estimation methodologies and evaluation models. To address this fundamental challenge in occluded posture recognition, we introduce a new thermal-based skeletal ground-truth derived from common motion-capture systems that models individual joints with visible thermal markers as shown in Figure 8.

As with common motion capture systems, this simple thermal marker system is designed from a standard form-fitting suit equipped with 9 solid nickel spheres with an approximate diameter of 3.0[cm] that match the pre-defined skeletal configuration. These solid metal spheres are attached to the suit at various locations that correspond to the joint positions of the patient. During the training process, these markers emulate the methodology of tracking joints by increasing their thermal intensity as an external heat source. Within our current prototype, we used passive heat sources such as briefly heating up the markers and then attaching them to the suit. An automated alternative is to provide an active heat source through a current source, however this method would incur additional safety precautions. This method provides a highly-accurate method for providing a ground-truth of the patient's posture, but also requires a manual configuration. The image provided in Figure 4 shows the simple design of the training suit with the attached solid nickel spheres used in the training process.

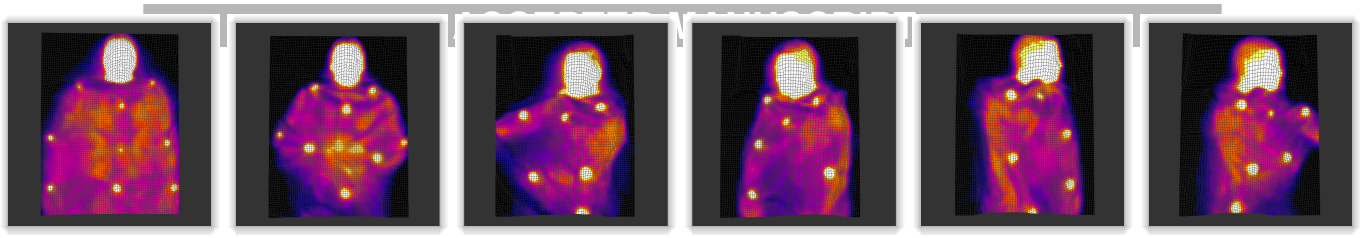

Figure 8: Thermal posture ground-truth. The image sequence illustrates the thermal ground-truth of the skeletal joints provided by the heat distribution of the suit conduction points. These fixed attachments irradiate heat that corresponds to the joint pattern required for the pre-defined skeletal joints and can identified in 3D space due to their position within the thermal point-cloud.

The result of the thermal skeletal ground-truth is the product of a simple adaptive thresholding and a connected-component algorithm that identifies the thermally intense regions of the spheres within the image. In the resulting thermal-cloud, the spheres appear as small white regions indicating the locations of the joint positions, as shown in Figure 8. For each grouping of points belonging to a joint, the unique joint position is calculated as the center of mass of this cluster. For labeling we employ a simple a semi-automated tool to assist in the identification of the skeletal joints for the training data. Adjacencies can then be generated for the required structure and can account for missing joints. Based on the provided adjacencies, the system will automatically generate the configuration required for the ground-truth posture.

6. Volumetric Thermal Modeling

Sleep-study occluded posture estimation offers a large reduction in both the degrees of freedom in both the patients movement and the volumetric region they occupy. Based on the assumption that the patient resides at rest within a limited spatial region and the occluding surface is covering the patient, this region of interest is easy to identify and model as a continuous enclosed thermal volume. This is achieved through the use of several assertions about the experimental setup: the patient resides within the bounded region and is supported by a rest surface, the occluding surface is supported by the patients body and does not penetrate through the volume of the body, the human body is contiguous, and the patient's face is visible and unobstructed. In this section we build on these assumptions to formulate the three-stage process of building the patient's posture volume and generating the associated volumetric model: (1) volume enclosure, (2) sphere hierarchy generation, and (3) the generation of a voxel grid that represents the thermal distribution of the patient's posture. This process and the resulting thermal distribution that models the thermal posture of the patient are shown in Figure 9.

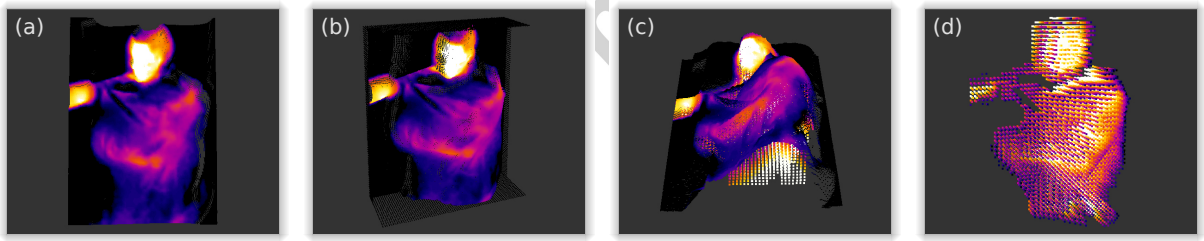

Figure 9: Volumetric thermal model process overview. The image in (a) illustrates the raw thermal cloud, (b) illustrates the enclosed region of this cloud, (c) illustrates the generated internal thermal distribution of the patient, and (d) provides the result of both the reconstruction and the thermal propagation through the enclosed volume, providing the thermal distribution in (d).

6.1. Posture Volume Enclosure

To begin the process of imposing constraints on joint locations within the occluded region, we enclose the volume between the recorded thermal-depth fusion surface and the known (depth-measured) background plane of the surface. Since the enclosed volume is a direct function of the occluded surface model and the bed surface, we assume that the contact surface of the bed can be obtained through a preliminary scan of the bed surface taken while patient is not present and the occluded surface model can be recorded in real-time. This encloses the volume of interest between these two surfaces (the bed surface and the occluding surface) that we populate with a scalar field generated through a volumetric sphere hierarchy.

6.2. Volumetric Sphere Hierarchy

To model the internal volume of the patient behind an occluded region, we introduce a simple and robust method for populating the area using discrete unit spheres through a methodology derived from simple *sphere-packing*. Generating this volume requires an enclosed region that is defined by the point-cloud data provided by the imaging devices included in the proposed prototype. From the enclosed region occupied by the patient defined by the beds surface and the recorded depth image, the volumetric reconstruction process used to define the occluded volume is derived from the 3D grid-based sphere-packing algorithm used to generate a *spherical hierarchy*.

This methodology is used as the basis of the volume reconstruction algorithm due to three assertions of the cloud that encapsulates volume of the patient: (1) the volume may be concave and contain complex internal structures and (2) the internal region may contain holes or regions that further reduce the patients potential joint positions due to volumes that are too small to occupy the associated joint, and (3) both surfaces sufficiently enclose the volume leaving no holes or gaps within the volumes surface. Sphere-packing is a simple algorithm that propagates unit spheres through a hollow region until some boundary conditions are met. This is based on three primary components commonly defined for sphere-packing: (1) the start position of the propagation, (2) the method of propagation, and (3) the boundary conditions must be defined for each sphere added to the volume. For (1), the starting position of the propagation is defined as the center of mass of the patients head. From our assertion that the patients head will always be uncovered, we can easily segment and identify the patients head within the thermal image due to the heat intensity of the patients face. The method of propagation (2) is derived from a bread-first search pattern. For the boundary conditions (3) of the propagation, we consider two primary boundaries: the point-cloud that encloses the region and regions that have very limited thermal intensities. This limits the propagation of the volume to regions that contribute to the patient's posture. Most importantly, the structure of the hierarchy provides context to how the volume is formed from the head joint. Shallow regions within the hierarchy represent joints that are closer to the head joint while deep regions within the hierarchy have a higher probability of representing limb or spine joints.

While this method provides an effective means of finding the continuous volumetric regions bound by complex surfaces, it has two main drawbacks: (1) the performance of this approach is equivalent to breadth-first search and does not scale well to a parallel algorithm, and (2) if there is any hole within the surface, the propagation could potentially diverge, which invalidates the data. To address these potential problems and introduce a close approximation of this algorithm we have also implemented a *grid-based* approach. This approach simply computes the scalar field between the two surfaces based on the dimensions of the original depth and thermal image sources. This technique is extremely simple and has a trivial parallel structure. This method also scales well to accommodate arbitrary body sizes within the reconstructed volume and does not impose a large additional overhead for larger body volumes.

6.3. Thermal Extended Gaussian Images (TEGI)

Extended Gaussian Images (EGIs) represent a mapping of surface normals of an object onto a unit sphere through a simple projection. This formulation provides an alternative form of representing complex geometric structures using a simplified form while maintaining the original geometric distribution. To reduce the resolution of the volumetric data provided by the thermal-cloud, we introduce the use of Thermal Extended Gaussian Images (TEGIs) to represent a projection of localized thermal intensities from the recorded thermal images onto the surfaces of the unit spheres within the sphere hierarchy.

TEGIs are introduced to establish a transfer function between the known recorded surface temperatures and the volumetric data represented by the sphere hierarchy within the occluded region. This function represents a conversion of the 2D thermal data residing within the surface lattice to a volumetric representation of the transferred heat and an estimate of the source direction. This allows the thermal data of the recorded surface point-cloud to be transfered to the newly generated internal volume that represents the patients potential posture constraints. Based on this model, TEGIs are used to represent both thermal intensity and directionality of the observed thermal distribution.

Each surface sphere within the hierarchy contains an TEGI that is parametrized by two characteristic features based on the on the sample points residing within the local neighborhood $(2r)$ of the sphere: (1) the thermal intensity t and (2) the Euclidean distance d between the contributing point and the sphere. This provides a parameterized distribution that models the local heat distribution across the surface of the recorded thermal cloud as a 2D Gaussian function $TEGI(t,d)$:

$$TEGI(t,d) = \alpha t e^{\left[-x^2/2(\beta d)\right]+\left[-y^2/2(\beta d)\right]} \tag{1}$$

Where the parametrization of the standard Gaussian distribution is defined by the thermal contribution t and scaled by a scalar thermal multiplier α provided by the thermal image. The distribution of the function is then modified by modeling σ^2 as the Euclidean distance between the point d and the center of the sphere with a distance scalar multiplier β where the value for the scalar multiplier β is defined by the device distance to the surface of the patient.

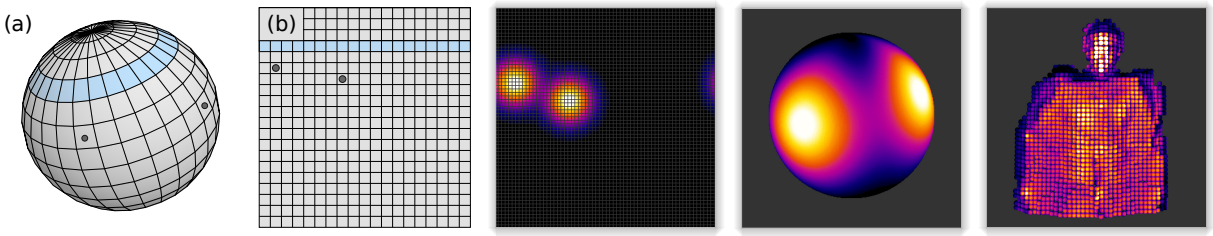

Figure 10: Extended Gaussian Image (EGI) spherical mapping [16] of thermal points to a volumetric region (a). For each thermal point within the recorded thermal point-cloud, the projection of the point will produce a location on the mapped unit sphere that will reside within a bounded surface region (b). The intensity of these projected thermal distributions then provide the volumetric representation of the heat transfer from the thermal point-cloud to the occluded volume.

The primary requirement of generating a TEGI is a procedure for projecting and mapping thermal points from the thermal cloud onto the surface of a unit sphere. To achieve this, a discrete form of the unit sphere is divided into discrete regions following the approach defined in [16] for automated point-cloud alignment. Then for each point within the local neighborhood, the point is projected onto the surface of the sphere and then assigned a 2D region index within the TEGI. This index will be used to identify the peak of the Gaussian distribution that will be added to the discrete surface representation of the sphere. Since the resolution of the Gaussian is discretized on the surface of the sphere, we sample the continuous parameterized Gaussian function at a fixed interval and allow the distributions to wrap around the surface of the sphere. The image in Figure 10 provides an illustration of how points are projected to the surface of a unit sphere and then used to generate the positions of the Gaussian distributions within the surface image of the sphere.

The contribution of multiple points within the same local neighborhood is accounted for through the addition of several different Gaussian distributions to the surface of the sphere, each with its own parameterization derived from its relative position to the sphere and its thermal intensity. The resulting TEGI is then defined as the sum of the contributions from all local points within the defined search radius. This defines the total thermal contribution of sphere $\mathcal{S}$ to the volume for the set of points within the spheres local neighborhood $\mathcal{N}$:

$$\mathcal{S}(p) = \sum_{i=0}^{n}\sum_{j=0}^{n} \alpha p_t e^{-x_i^2/2(\beta d)+-y_j^2/2(\beta d)}, \quad \forall p \in \mathcal{N} \tag{2}$$

Geometrically, the contribution of each points thermal intensity to the surface of the sphere also incorporates the directionality of the thermal intensity of the point in the direction of the sphere. This provides a rough estimate as to the direction of the source of the thermal reading identified at the surface point. While this approximation of the heat transfer function does not provide an accurate model of the inverse heat transfer problem, it provides an effective means for estimating the inverse propagation of the heat measured at the recorded depth-surface to define the thermal signature of the volume.

These TEGIs are then evaluated for each sphere in the spherical hierarchy that reside within the surface of the thermal cloud. The resulting thermal intensity of each sphere is then used as the seed for propagating the observed heat through the patient's posture volume. These thermal values are then used generate a three-dimensional voxel model of the patients heat distribution.

6.4. Thermal Voxel Grids

To integrate the thermal contribution of each TEGI within the constructed sphere hierarchy, the grid-based nature of the propagation algorithm is used to populate a scalar field of the thermal values into a voxel grid. This voxel grid provides the thermal distribution of the internal volume of the patient used to represent the thermal distribution unique to a specific posture. This distribution is then used to represent the patient's posture as a 3D image that can be classified based on a pre-trained set of postures that contain associated skeletal joint positions. The ground-truth skeletal joint positions, volumetric reconstruction of the patient's posture, and the resulting skeletal posture classification is generated from this method are shown in Figure 11. While the accuracy of some skeletal configurations are limited by the ground-truth skeletal joint locations, we still obtain reasonable classifications of the joint positions for posture instances that have completely occluded joints when the training suit is not included. Therefore, we validate the use of the thermal distribution as a classifier for occluded skeletal posture estimation.

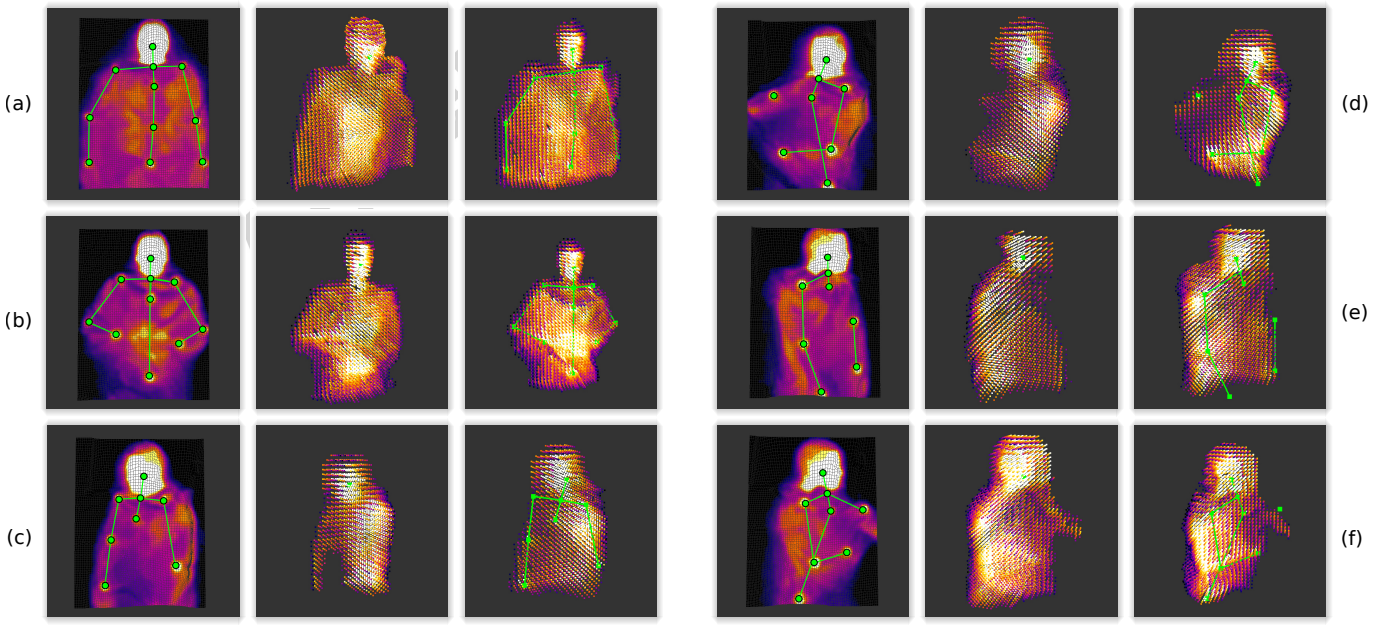

Figure 11: Skeletal posture estimation results for six standard sleeping postures. The first image in each sequence provides the ground-truth skeletal posture with superimposed skeletal joints, followed by the middle image that illustrates the thermal distribution used to obtain the trained skeletal posture rendered in the last image of each sequence. This process illustrates both the volumetric reconstruction and the posture classification result used to identify the six postures.

7. Thermal Skeletal Volumetric Training

The underlying correlation between volumetric thermal distributions and skeletal joint positions used to formulate our posture estimation is defined by two primary factors: (1) the skeletal ground-truth of a patients posture and (2) the thermal distribution of the patients volume within the occluded region. Together, these two components form the training and identification data used to estimate the occluded skeletal posture of the patient within an occluded region. There are several types of training methodologies and models that have been designed for three-dimensional medical image classification. Of these methods, Convolutional Neural Network (CNNs) [17] and Deep Neural Networks (DNNs) [17] are the most commonly used methods for identifying complex structures within 3D images. In the proposed method, we have selected a feed-forward CNN-based network structure to handle the higher dimensionality of the 3D thermal voxel grid we generated within Section 6. This is due to the dense representation of the patient's thermal distribution rather than a feature-based estimation which would better suit a DNN-based method. Therefore we allow the CNN to generate features through sequential filters that identify thermal-specific classification metrics. In our method we implement CNN with 4 fully-connected layers with rectified linear units (ReLUs) which obtain results faster than traditional *tanh* units [18]. Additionally, since there is no analytical method to determine the optimal number of convolutional layers for a given application, our network structure is determined empirically based on the correct identification of posture states.

Constructing Training Models. Based on the three-dimensional representation of our training set that corresponds to a set of ground-truth skeletal estimates, we construct the training models of our posture estimation based on individual thermal distributions and their associated skeletal joint positions. Our training model relies on the construction of these components for each posture as they are provided to our CNN implementation. Therefore to incorporate several individual postures, we repeat the reconstruction and the skeletal ground-truth process for the most plausible postures that can be obtained within our limited spatial domain to provide an adequate recognition set. From this set of potential postures, each dense thermal distribution can be used to represent a thermal distribution that identifies the skeletal configuration and joint positions that are retrieved as part of the classification result. However, this distribution does not uniquely identify this individual posture. This is due to several other factors that may lead to a different thermal distribution representing the same posture including differences in the heat distribution, the depth of the occluding surface, and number of layers between the patient and the thermal camera. In an attempt to address this problem and provide a reliable training set we record several variants of the same postures with slight variations within the recorded thermal distribution. The objective of constructing the training models using this method is to address variance in body temperature, distribution, and material coverage of the patient.

Neural Network Structure. For a given dense thermal distribution we formulate the feature value at location (i, j, h) in the k-th 3D feature map of l-th layer, $z^l_{i,j,h,k}$ is calculated by $z^l_{i,j,h,k} = w^{l^T}_k x^l_{i,j,k} + b^l_k$, where w^l_k and b^l_k are the weight vector and bias term of the k-th filter of the l-th layer respectively, and $x^l_{i,j}$ is the input patch centered at location (i, j) of the l-th layer. The kernel w^l_k that generates the feature map $z^l_{:,:,:,k}$ is shared. This weight sharing mechanism can reduce the model complexity and make the network easier to train for our dense volumetric model. We utilize standard activation function (ReLUs), which are desirable for multi-layer networks to detect nonlinear features that may arise within our thermal volume. We let α denote the nonlinear (ReLUs) activation function. The activation value $\alpha^l_{i,j,k}$ of convolution feature $z^l_{i,j,k}$ can be computed as $\alpha^l_{i,j,k} = \alpha(z^l_{i,j,k})$. The pooling layer is used to create the shift-invariance by reducing the resolution of the feature maps that correlate to the thermal distribution of the posture. Each feature map of a pooling layer is connected to its corresponding feature map of the preceding convolutional layer. We let p be the pooling function, for each 3D feature map $\alpha^l_{:,:,:,k} : y^l_{i,j,hk} = p(\alpha^l_{m,n,q,k}), \forall (m,n,q) \in R_{ijh}$. In that, $R_{i,j,h}$ is a local neighborhood around location (i, j, h) in the voxel grid. The kernels in the 1st convolutional layer are designed to detect the edges and curves of the image generated by the thermal propagation algorithm. The higher layers learn to encode more abstract features of the thermal distribution, allowing us to consolidate several convolutional and pooling layers together, to extract higher-level feature representations. In order to train our model, we try to minimize the error associated with handling a variety of training sets that have minimal distinguishing factors between volumetric models. We define N desired input-output relations $x^{(n)}, y^{(n)}; n \in [1, ..., N]$, where $x^{(n)}$ is the $n-$th input data, $y^{(n)}$ is its corresponding target label (the posture classification) and $o^{(n)}$ is the output of the classification. The loss of CNN can be calculated as follows: $L = \frac{1}{N} \sum_{n=1}^{N} l(\theta; y^{(n),o^{(n)}})$. By minimizing the loss function, we can find the best fitting set of parameters that allow us to maximize cross-patient features to improve the reliability of retargeting existing training sets to new patients.

Learning Model Implementation. We trained our classification network to detect 6 postures of the patient based on our generated thermal voxel grid images. The classification label (one of six postures) is assigned for each thermal distribution. 60 thermal voxel grid images are used for training while 180 other distributions have been used for testing. We avoid overfitting through two common methods: First, we apply *Dropout* to randomly drop units (along with their connections) from the neural network during training [19], which prevents neurons from co-adapting. Second, cross-correlation is applied to stop the training when the cross-validation error starts to increase, leading to our termination condition. Additional convolutional layers generally yield better performance but as the performance gain is reduced, we see diminishing returns in the training process. Therefore the number of connected layers required to avoid overfitting is commonly defined as two as referred in [17]. We also applied early stopping mechanism to make sure that the learning process is terminated when the network reaches to a certain status. The criteria to stop the training is as following: (a) the training classification success rate has reached a sufficient classification percentage, (b) the learning process reaches 500 iterations, or (c) the cross-entropy is lower than 0.005. This mechanism prevents the classifier from memorizing counter-productive features from some samples, especially within low quality datasets.

8. Experimental Results

Driving the experimental results of the proposed volumetric model for skeletal posture estimation, we identified several common sleep postures that exhibit a wide variety of skeletal joint positions that form both partial and complete posture estimates due to the visual occlusions introduced by the use of a standard blanket. Based on these common postures, our objective is to collect the skeletal ground-truth, generate the associated thermal distribution, and then correlate this distribution with the recorded skeletal joint positions for the patient's training set. From the generated training set, we can then estimate the patient's approximate skeletal posture solely based on their current thermal distribution. This process is used to identify several standard sleep postures and their associated skeletal joint positions for a patient visually occluded by a blanket within a standard bed. During these experiments we tested with different standard blanket types (thin, thick, plush), and explored the accuracy of our training set based on two criteria: (1) classification of an individuals posture using their own training set and (2) using an extended training set to classify the posture of other individuals.

Standard Posture Estimation. The primary qualitative metric for both identifying a patient's posture and associated skeletal structure in occluded regions is based on the ability to recognize the posture and the accuracy of the generated skeletal joints used to represent the patient. In these experimental results, we perform a quantitative analysis for the accuracy of the of this method with respect to identifying the correct posture based on the generated thermal distribution. The image sequences in Figure 11 (a-f) illustrate six common postures along with their associated ground-truth skeletal measurements as the first image within each sequence. The posture sequence for these experiments is defined as: (a) face up + arms at the side, (b) face up + hands on chest, (c) face left + straight arms, (d) face left + bent arms, (e) face right + straight arms, and (f) face right + bent arms. The second image within each sequence provides the rendered thermal distribution of the patient based on the voxel data generated from the volumetric model. This data is then used to identify the associated skeletal structure, as presented in the last image of each sequence. The images shown within Figure 11 correlate to the numerical classifications presented within Figures 12 and 13.

Individualized Posture Estimation. As the primary quantitative metric of the volumetric distribution method, we measure the accuracy of the classification of the patient's posture based on our six standard postures. For each posture, we collect the ground-truth and 40 variants (with subtle movements) to provide a sufficient training set applicable to the limited spatial domain within which we can define a discrete posture set. This results in 240 data sets in total, with 60 used for training and 180 data sets utilized for testing. The confusion matrix illustrated in Figure 12 shows the performance of the classification rate for the trained system. This process was conducted using the training data of a single individual (heigh: 6[ft], weight: 150[lbs]), resulting in an average ~94.45% classification accuracy for identifying an individuals posture based on their training set.

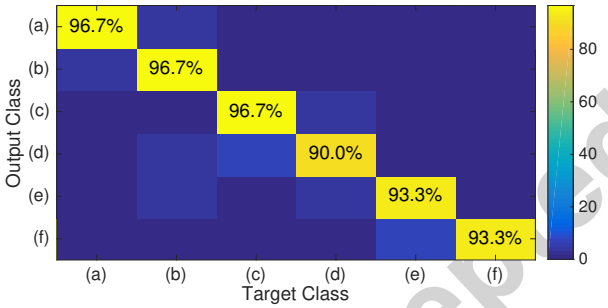

Figure 12: Individualized confusion matrix for the six identified postures. The correlation between the postures (as shown in Figure 11), illustrates a ≈ 90% classification accuracy. Similar postures incur misclassification due to changes in the patient's joint locations.

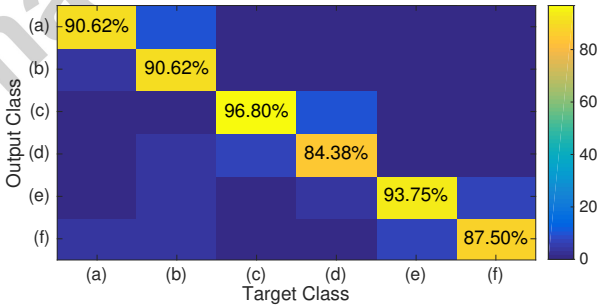

Figure 13: Confusion matrix illustrating the accuracy of the estimation using a set of multiple patients that did not contribute to the training of the CNN used to perform the classification, this mitigates the requirement of per-patient training for similar body-types.

Cross-patient Posture Estimation. Individual body structure plays a significant role within posture estimation algorithms that do not use features, however based on the generalized volumetric model of the body used to classify the identified skeletal posture, this method can also be loosely applied across several patients with similar body volumes, obtaining reasonable results. The confusion matrix in Figure 13 shows the classification results of the postures provided by three individuals based on a pre-trained posture set formed from a single individual, with an average accuracy ~90.62% for identifying a secondary individuals posture based on the mismatched training set. This accuracy is due to participants similar body volume characteristics, which are primarily defined by height and weight. All subjects varied from 5[ft] - 6[ft] and weigh 130[lbs] - 180[lbs]. To improve the accuracy for other body volumes, additional training sets based on weight and height could be implemented (in the same way as standard depth-based tracking methods).

Impact of Training Network Structure. The introduction of additional layers within the CNN improves the performance of classification in both experiments, but we still observe diminishing returns. We tested the CNN from 1 to 4 convolutional layers. The results are as following. (1) With 1 layer, the accuracy obtained up to 77%, with 1.3 millions # of weights, the training time is approximate 5 minutes. With 2 layers, the accuracy could go up to 88% with around 2 millions weights and 10 minutes of training. When the number of layers goes from 3 to 4, it will takes around 15, 20 minutes, with accuracy of 92% to 96% using 2.8 to 3.2 millions number of weights, respectively.

9. Discussion

There are three primary considerations employed within the design of these results that are addressed within the current version of this approach: (1) the training set is based on a discrete enumeration of skeletal postures, limiting the skeletal movement resolution, (2) the entire voxel volume is utilized with CNN generated features, so training is based on body volume size, and (3) skeletal refinement algorithms for fine-grain position and orientation estimates have not been employed, thus the resulting skeletal movement between enumerated training postures is discrete. These issues can be properly addressed through providing an extensive training set of postures from numerous patients (as common with all depth-based skeletal tracking), feature localization and extraction, and a joint refinement algorithm that compensates for the disparity between the trained skeletal structure and the patient's actual joint positions, all of which are extensions of this method described within this section. The implemented volumetric reconstruction algorithm also provides a means of accurately modeling and visualizing the volumetric posture of a patient within an occluded region without accurate joint estimations. This allows the this method to be applied to numerous additional medical imaging applications such as patient monitoring and thermal distribution modeling for other various studies.

Ground-truth Tracking. The introduction of the thermal-based motion tracking suit provides an effective means of identifying joint positions through the occlusion medium, but also requires a laborious setup time. While this process is only required for the initial training and provides the only way of generating a ground-truth estimate for this method, it still requires additional effort to heat and trace training images. Within our implementation we chose to use an external thermal source to heat the metal ball joints due to the simplicity and safety of the system. Although due to the ambient environment temperature, these joint markers naturally dissipate heat rapidly. Therefore, an additional current-based system could induce heat through the joint markers to provide a more consistent heat source during the training process. This would greatly reduce the time and effort required to record thermal marker positions for skeletal joint positions. The only concern with this approach is that this would require generating heat from a voltage source that may pose potential safety risks due to current or the overall thermal intensity of the markers under an occluding medium such as a blanket where markers may overheat or become inconvenient to the patient.

Fine-grain Joint Position Estimation. The proposed skeletal estimation algorithm provides a coarse-grained estimate of the joint positions according to the reconstructed thermal distribution. This method allows us to accurately classify postures for sleep studies and provide detailed volumetric images of the patients body, but does not provide fine-grained joint orientation information. To address this, the introduction of thermal-depth features can be used to improve fine-grained position estimations of joints. The challenge imposed by using this standard form of feature classification is that within the thermal distribution, there remain a significant number of ambiguous cases due to the thermal variance in the surface material and unreliable depth features that do not correspond to joint positions. However, our proposed direction can be extended for the intersection between these two approaches by: (1) using the volumetric model to greatly reduce the search space for potential joint positions and (2) utilizing thermal-depth features to estimate fine-grained joint positions within this limited spatial domain. As in prior techniques feature-based techniques [4], per-pixel features can be created to identify class association within the thermal point-cloud.

Extended Applications. The techniques proposed within our work can also potentially apply to other domains such as obstructive sleep apnea [20, 21], and breathing disorder detection by monitoring the fluctuation in the reconstructed thermal volume as it corresponds to breathing behavior. The thermal-depth volumetric visualization technique we introduce can also be used to monitor patient comfort and could be used to identify limb temperature in patients that have limited mobility, which is a critical indicator of adequate circulation through the vascular system. This technique can also be used to help monitor sleep postures of post-surgical patients (ulcer/sore) [22] to ensure both proper circulation and that proper movement regiments are maintained for long-term patients. Additional applications that utilize the volumetric component of the thermal-depth fusion can also be used in several other domains for heat-flow visualization and non-destructive analysis.

Future Work. Improvement of the skeletal estimates through fine-grain tracking, larger training sets, and an improved accuracy of training body-type specific patients will address the challenges introduced in the current prototype system. In addition to the small data-set, CNN overfitting, and errors in joint estimation based on body shape, the current training methods need to be streamlined and expanded to incorporate additional body-independent features. In the current state, the proposed method and ground-truth technique show promise for volumetric thermal modeling, but require extended training and clinical evaluation. Based on the core proposed method of thermal-depth fusion, additional data-sets will assist with accuracy and reliability.

10. Conclusion

In this work we have introduced a novel approach for integrating thermal and depth imaging to form a volumetric representation of a patient's posture to provide occluded skeletal joint estimates for pre-trained sleeping postures. We have built on similar existing thermal-depth fusion techniques to provide a reliable thermal volume reconstruction process that provides an accurate high-resolution visualization of a patient's posture. By extending this approach to define a patient's unique thermal distribution, we have introduced a new method for correlating a patient's unique heat signature with our motion-capture inspired ground-truth estimate of the patient's skeletal posture for generating occluded joint positions. This result is illustrated through the application of our approach to six pre-defined postures with an average classification accuracy of ~94.45% for an individual and an accuracy of ~90.62% for the use of a trained network used for cross-patient posture esimtaitons.

References

[1] N. Mohsin, X. Liu, and S. Payandeh, "Signal processing techniques for natural sleep posture estimation using depth data," in *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Oct. 2016, pp. 1–8.

[2] M. Ye, X. Wang, R. Yang, L. Ren, and M. Pollefeys, "Accurate 3d pose estimation from a single depth image," in *2011 International Conference on Computer Vision*, Nov. 2011, pp. 731–738.

[3] J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, and A. Blake, "Efficient Human Pose Estimation from Single Depth Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2821–2840, Dec. 2013.

[4] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *CVPR 2011*, Jun. 2011, pp. 1297–1304.

[5] F. Achilles, A.-E. Ichim, H. Coskun, F. Tombari, S. Noachtar, and N. Navab, "Patient MoCap: Human Pose Estimation Under Blanket Occlusion for Hospital Monitoring Applications," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2016*, ser. Lecture Notes in Computer Science. Springer, Cham, Oct. 2016, pp. 491–499.

[6] X. Liu and S. Payandeh, "Toward study of features associated with natural sleep posture using a depth sensor," in *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, May 2016, pp. 1–6.

[7] P. Nguyen, X. Zhang, A. Halbower, and T. Vu, "Continuous and fine-grained breathing volume monitoring from afar using wireless signals," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, April 2016, pp. 1–9.

[8] P. Nguyen, S. Transue, M.-H. Choi, A. C. Halbower, and T. Vu, "Wikispiro: Non-contact respiration volume monitoring during sleep," in *Proceedings of the Eighth Wireless of the Students, by the Students, and for the Students Workshop*, ser. S3. New York, NY, USA: ACM, 2016, pp. 27–29.

[9] S. Transue, P. Nguyen, T. Vu, and M. H. Choi, "Real-time tidal volume estimation using iso-surface reconstruction," in *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, June 2016, pp. 209–218.

[10] ——, "Thermal-depth fusion for occluded body skeletal posture estimation," in *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE'17)*, July 2017, pp. 167–176.

[11] J. van Baar, P. Beardsley, M. Pollefeys, and M. Gross, "Sensor fusion for depth estimation, including tof and thermal sensors," in *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, Oct 2012, pp. 472–478.

[12] S. Iwasawa, K. Ebihara, J. Ohya, and S. Morishima, "Real-time human posture estimation using monocular thermal images," in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, Apr. 1998, pp. 492–497.

[13] I. Riaz, Jingchun Piao, and Hyunchul Shin, "Human Detection by Using Centrist Features for Thermal Images," *IADIS International Journal on Computer Science & Information Systems*, vol. 8, no. 2, pp. 1–11, Jul. 2013.

[14] K. Yasuda, T. Naemura, and H. Harashima, "Thermo-key: human region segmentation from video," *IEEE Computer Graphics and Applications*, vol. 24, no. 1, pp. 26–30, Jan. 2004.

[15] J. Zhang, W. Wang, and C. Shen, "Improved Human Detection and Classification in Thermal Surveillance Systems," Sep. 2010.

[16] A. Makadia, A. Patterson, and K. Daniilidis, "Fully Automatic Registration of 3d Point Clouds," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, Jun. 2006, pp. 1297–1304.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017. [Online]. Available: http://doi.acm.org/10.1145/3065386

[18] N. D. Lane, P. Georgiev, and L. Qendro, "DeepEar: Robust Smartphone Audio Sensing in Unconstrained Acoustic Environments Using Deep Learning," ser. UbiComp '15. ACM, 2015, pp. 283–294.

[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.

[20] A. M. Neill, S. M. Angus, D. Sajkov, and R. D. McEvoy, "Effects of sleep posture on upper airway stability in patients with obstructive sleep apnea." *American Journal of Respiratory and Critical Care Medicine*, vol. 155, no. 1, pp. 199–204, Jan. 1997.

[21] R. D. Cartwright, "Effect of sleep position on sleep apnea severity," *Sleep*, vol. 7, no. 2, pp. 110–114, 1984.

[22] K. Vanderwee, M. H. F. Grypdonck, D. De Bacquer, and T. Defloor, "Effectiveness of turning with unequal time intervals on the incidence of pressure ulcer lesions," *Journal of Advanced Nursing*, vol. 57, no. 1, pp. 59–68, Jan. 2007.