Thermal-Depth Fusion for Occluded Body Skeletal Posture Estimation

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

Department of Computer Science and Engineering, University of Colorado Denver

{shane.transue, phuc.v.nguyen, tam.vu, min.choi} @ucdenver.edu

Abstract—**Reliable occluded skeletal posture estimation is a fundamentally challenging problem for vision-based monitoring techniques. This is due to several imaging related challenges introduced by existing depth-based pose estimation techniques that fail to provide accurate joint position estimates when the line of sight between the imaging device and the patient is obscured by an occluding material. 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. 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 correctly identifying common sleep postures that present challenging cases for current skeletal joint estimations, obtaining an average classification accuracy of ~94.45%.**

I. INTRODUCTION

Accurate and reliable occluded skeletal posture estimation presents an interesting challenge for vision-based methods that heavily rely on depth-imaging [1], [2] to form accurate skeletal joint estimations [3], [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-based studies where patients are heavily occluded by both clothing and common forms of bedding. The clinical significance of introducing an occluded skeletal posture estimation that handles these cases is illustrated through the lack of accurate estimate solutions used for wireless tidal volume estimation [5] with clothing occlusions, long-term sleep-based respiratory monitoring [6], [7], [8] with bedding occlusions.

While recent depth-based imaging methods [9], [10] have begun exploring how to solve this problem, 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. 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 completely fail to identify obscured skeletal joints. In this work, we explore
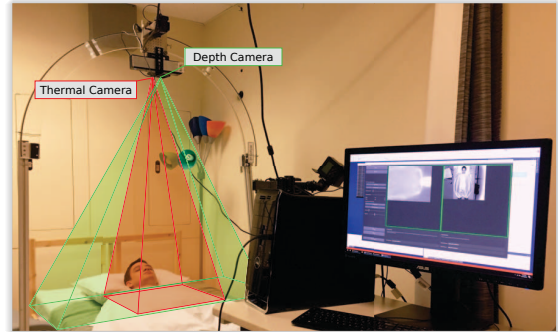
Fig. 1. Experimental setup for detecting occluded skeletal joints that define a patient's posture with occlusions from standard bedding. Emphasis: The proposed thermal-depth fusion skeletal estimation prototype that generates and reconstruct the 3D thermal distribution of the patient's occluded posture.

how the limitations of depth images can be supplemented through integrating thermal imaging to introduce new methods in occluded skeletal posture estimation.

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, we develop a real-time posture estimation derived from both depth and thermal imaging with the objective of providing a reliable means of estimating occluded skeletal postures.

In this novel approach to occluded skeletal posture estimation, we introduce three primary contributions: (1) we present a thermal-based marker system for obtaining an occluded skeletal ground-truth posture estimate derived from modern motion capture techniques for defining occluded skeletal joint positions, (2) develop a volumetric representation of patient's thermal distribution within an occluded region, and (3) introduce a coarse-grained skeletal posture estimation technique for identifying joint positions of visually obscured patients. 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 a patient's completely occluded joint positions.

II. 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 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.

Depth-based Skeletal Estimation. The pioneer work for depth-based skeletal estimation from a single depth image for the Microsoft Kinect devices [3], [4] 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.
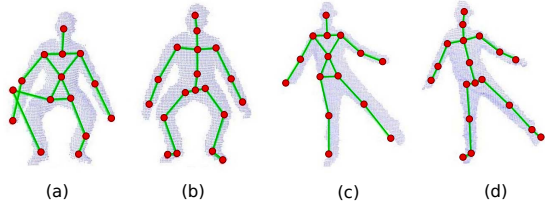

Fig. 2. Skeletal posture estimations from recent techniques from the Microsoft Kinect, Primesense OpenNI (a, c), and improvements (b, d) by [3] that utilize depth-imaging to accurately identify joint positions in non-occluded applications. These methods have been further refined and extended with the introduction of newer depth-imaging devices such as the Microsoft Kinect2.

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 [9]. 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 [10] 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.

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. Early work presented in [11] developed a simple algorithm for detecting the skeletal structure within a two-dimensional image, but the applications of this method are limited and cannot be utilized to form a 3D 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 [12] and tracking generalized human behaviors [13], [14] 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. 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.

III. 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. 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. In this section we enumerate several primary challenges associated with thermal imaging that greatly complicate thermal-based skeletal estimation.

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 RGB 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 instances difficult or completely intangible.
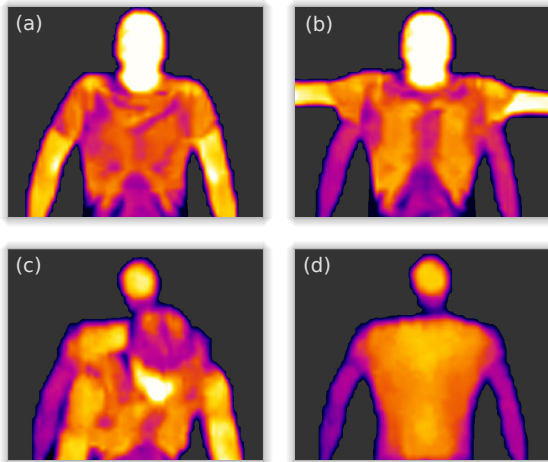

Fig. 3. Skeletal posture estimation challenges associated with thermal imaging. The image in (a) illustrates an ideal non-occluded thermal image but illustrates non-uniform thermal distribution of a patient's thermal signature, (b) provides an illustration of heat marks left by a patient's arm movements, (c) illustrates thermal ambiguities of the patient during motion, and (d) illustrates the patient's residual heat left when the patient has been removed.

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 patient-based 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. Additionally, with the introduction of an occluding material *tent-effect*, limb position may contribute to a 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.

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 in the recorded thermal image. 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.

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

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 posture 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 lead to ambiguities 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.

IV. 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 III. 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.

A. 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.
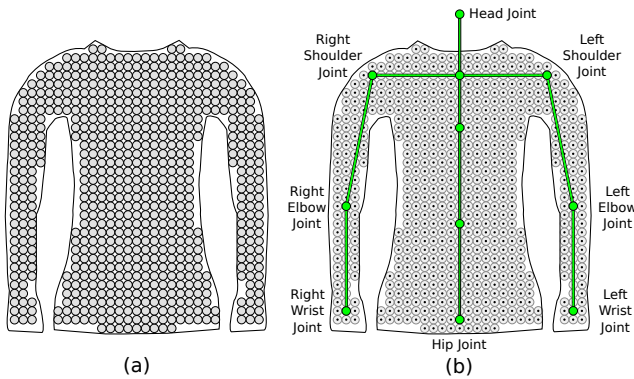
Fig. 4. Volumetric reconstruction of an ideal skeletal posture. The image in (a) illustrates a discrete approximation of the patient's volume. The image in (b) provides an illustration of the mapping between a voxel representation (black dots) of this volumetric data and the ground-truth skeletal estimate of the posture (illustrated as a set of joints and associated bones).
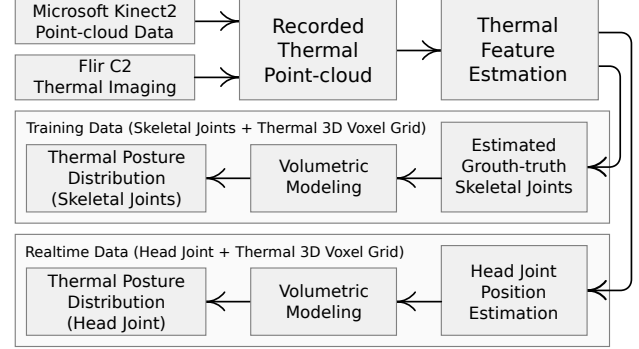
Fig. 5. Overview of the proposed approach for reconstructing the volumetric thermal data that contributes to the occluded skeletal posture estimation. This includes the generation of the volumetric data with the skeletal ground-truth for training and the real-time data with the provided head joint used during the occluded posture estimation process.

This 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 that can be assigned to an associated skeletal estimates that may contain visually ambiguous joint positions through training.

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, meaning that they are not visible, but reside within this volume. Due to these commonly occurring conditions that are not well handled by existing methods, the proposed method is based on creating a correlation between the patient's volumetric thermal distribution and an associated skeletal posture. Based on this correlation, if the known skeletal joint positions are provided for the observed thermal distribution, we can estimate the patient's skeletal posture even when the subject is highly occluded, has several ambiguous joint positions, or the skeletal components are disconnected.

B. 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 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 thermal posture signature for a patient is defined below:

1) Thermal Cloud Generation (Depth + Thermal)
2) Patient Volume Reconstruction (Sphere-packing)
3) Surface Heat Propagation (Extended Gaussian Images)
4) Volumetric Heat Distribution (Thermal Voxel Grid)

This 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. This forms two different tracks within the core algorithm of our approach which are defined within the data-flow of our technique presented in Figure 5.

V. 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.

A. 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 to an image size of 320x240. 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 to generate the thermal-cloud of the volume enclosing the patient due to the occluding material.
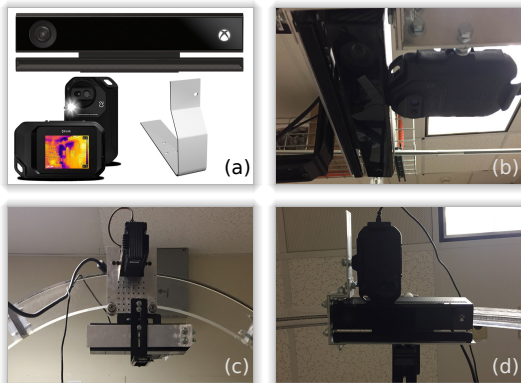

Fig. 6. Thermal posture device prototype. The two devices (Kinect2, C2) are mounted with a fixed alignment provided by the bracket shown in (a). The images in (b-d) illustrate the mount attached to the bed rail with both devices.

Fig. 7. Thermal posture ground-truth and training suit (a) with attachable metal spheres (b). The suit is worn during the training process to identify the relationship between the patient's thermal volume and joint positions (c).

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.

B. 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 [9], this remains a significant challenge in occluded posture estimation methodologies and evaluation models.

To address this challenge we introduce a new thermal-based skeletal ground-truth derived from common motion-capture systems. 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]. 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 known joint positions. This provides a highly-accurate method for providing a ground-truth of the patient's posture while an
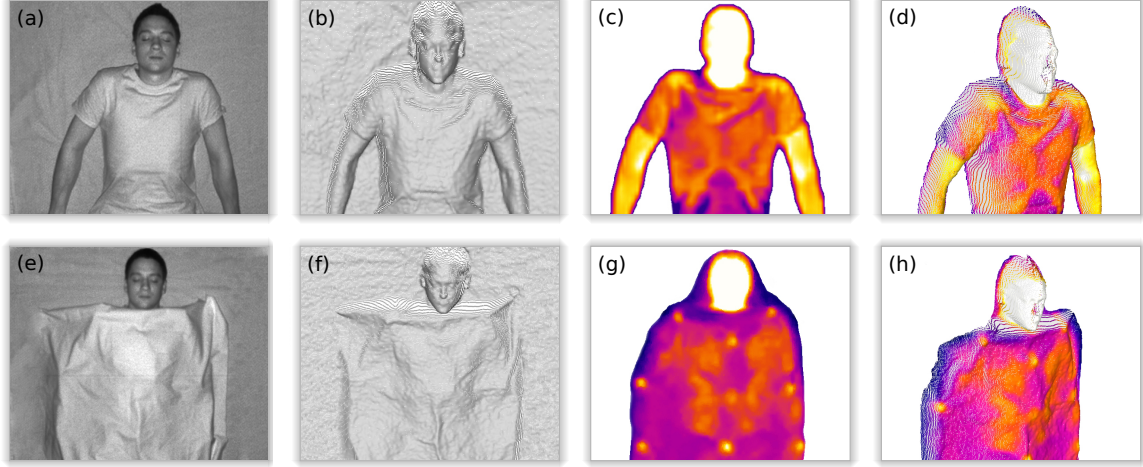
Fig. 8. Thermal surface point-cloud acquisition. The sequence of images illustrate the data collected from both the Microsoft Kinect2 and Flir C2 thermal devices to obtain thermal and surface point-cloud data. The images (a-d) illustrate the collection of the infrared, depth, thermal, and thermal surface respectively for a non-obscured view of the patient. The images (e-h) illustrate this data sequence for the same supine skeletal posture with an occlusion material present. Surface details provided by depth imaging (f) fail to provide a reliable means of estimating skeletal joints (for example, identifying hand joint positions in (e-f) is extremely difficult). Using the proposed ground-truth estimation, we can assert known joint positions through occluding materials.

occluding surface is present. The image provided in Figure 7 illustrates the simple design of the training suit with the attached solid nickel spheres used in the training process.

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 (g). 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. Based on the provided adjacencies, the system will automatically generate the required skeleton. For occluded joints, we introduce a *partial skeletal structure* (Figure 9).
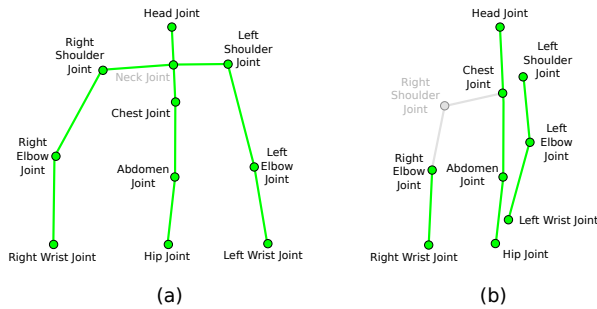

Fig. 9. Thermal skeleton ground-truth. The ground-truth skeleton presented in (a) illustrates a complete skeletal posture based on every supported joint being identified. The skeleton presented in (b) represents the patient in a left facing posture with the right shoulder joint completely occluded.

The disconnected skeletal structure we provide presents a best-case posture estimate based on the provided thermal information within the model. This allows us to provide a partial solution for instances where the occluding material may prevent several joints from being recognized in both thermal and depth images, for which we have no obtainable solution.

VI. 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 region and the occluding surface is covering the patient, this region of interest is easy to identify and model as a continuous enclosed volume as illustrated in Figure 8 (f). 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.

A. Posture Volume Enclosure

To begin the process of imposing constraints on the possible joint locations within the occluded region, we enclose the volume between the recorded depth image and the known bed surface. Since the enclosed volume is a direct function of the occluded surface model provided by the point-cloud and the bed surface, we assume that the contact surface of the bed can be obtained by a simple planar model or through a preliminary scan of the bed surface taken while patient is not present.

B. 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 two 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.
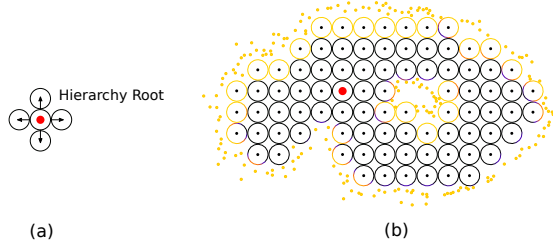

Fig. 10. Two-dimensional variant of the volumetric reconstruction algorithm. The image in (a) illustrates the hierarchy root and the propagation directions and (b) illustrates the limitation of the propagation by the surrounding point-cloud and associated thermal intensities of the depth points.

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. The image in Figure 10 illustrates this thermal 2D sphere-packing algorithm. For our three-dimensional skeletal posture data, the root position resides within the head of the patient.

C. 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 representation. 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.
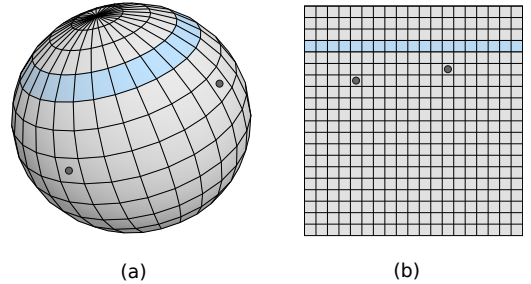

Fig. 12. Extended Gaussian Image (EGI) spherical mapping [15]. For each thermal point within the recorded thermal point-cloud, the projection of the point will produce a location on the unit sphere that will reside within a bounded surface region. These surface regions are defined by the height and width of the EGI map (b). The corresponding surface regions in (a) are displayed in the two-dimensional representation in (b).

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 [15] for automated point-cloud alignment. Then for each point within the local neighborhood, the point is projected onto the surface of the
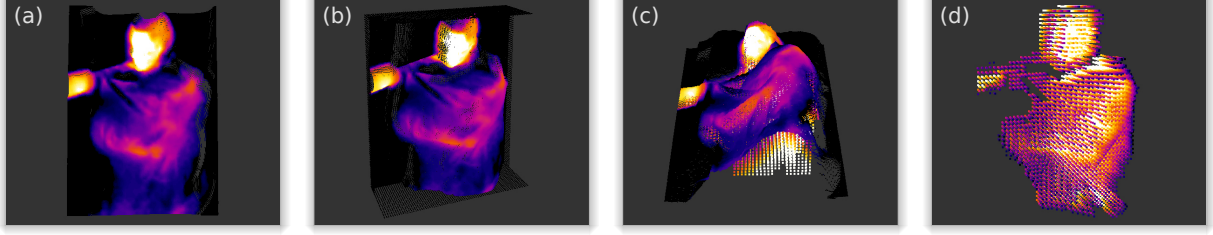
Fig. 11. Volumetric thermal model process overview. The image in (a) illustrates the raw thermal cloud, (b) illustrates the enclosed region of this cloud, (c) illustrates the generated internal thermal distribution of the patient, and (d) provides the result of both the reconstruction and the thermal propagation through the enclosed volume. The thermal distribution in (d) will then be provided to the training algorithm with an associated skeletal estimation.

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 12 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 $\mathcal{N}$:

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

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.
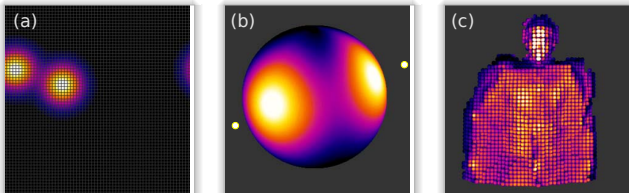

Fig. 13. Thermal Extended Gaussian Images for the distribution of heat due to surrounding thermal points. The image in (a) represents the discrete TEGI map of the sphere surface that contains the thermal contribution of two points, (b) illustrates the TEGI in 3D space with the two contributing points, and (c) provides a rendering of the TEGIs within the sphere hierarchy used to show the thermal propagation from the surface scan.

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.

D. Thermal Voxel Grids

To integrate the thermal contribution of each TEGI within the constructed sphere hierarchy, the grid-based nature of the propagation algorithm used to generate the volume is used to populate a scalar field of the thermal values into a voxel grid. This fixed-dimension voxel grid provides the thermal distribution of the internal volume of the patient used to represent the thermal distribution of a unique posture. The thermal distribution residing within the voxel grid 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. An example of the resulting 3D image illustrating the patient's posture within the voxel grid is illustrated in Figure 14 (d).

VII. 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.

Neural Network Structure. 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) [16] and Deep Neural Networks (DNNs) [16] are 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 VI. 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
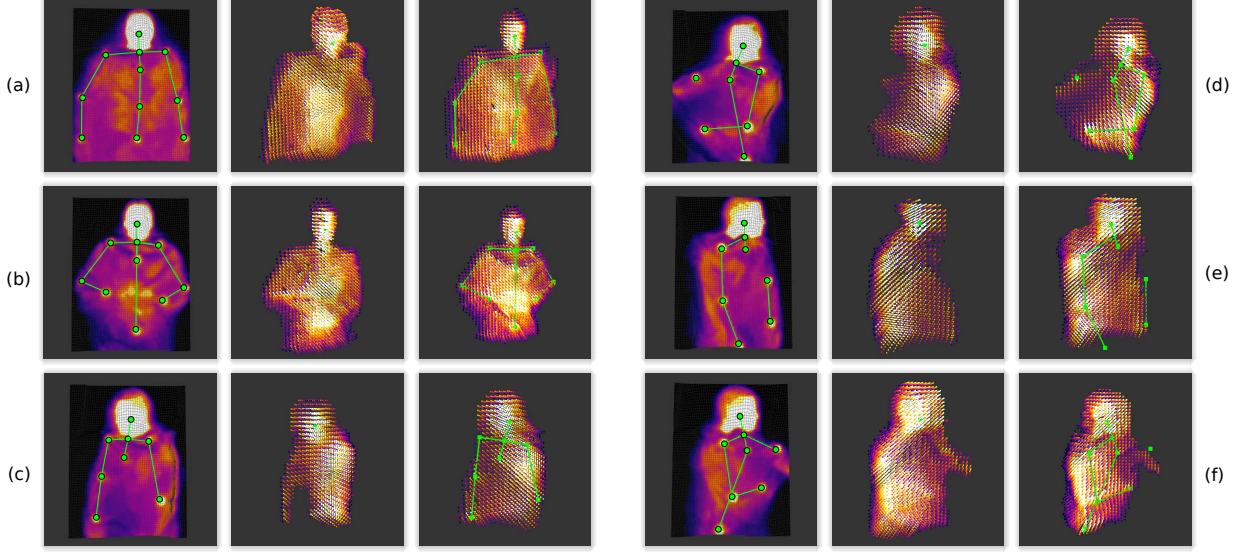
Fig. 14. Skeletal posture estimation results for six standard sleeping postures. The first image in each sequence provides the ground-truth skeletal posture, followed by the middle image that illustrates the thermal distribution used to obtain the trained skeletal posture rendered in the last image of each sequence.

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 [17]. 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.

Training Model. We trained CNNs 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 [18], 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 [16].

VIII. 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.

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 14 (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.

Individualized Posture Estimation. As the primary quantitative metric of the volumetric distribution method, we measure the accuracy of the classification of the patients 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 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 15 shows the performance of the classification rate for the trained system, resulting in an average ~94.45% classification accuracy.
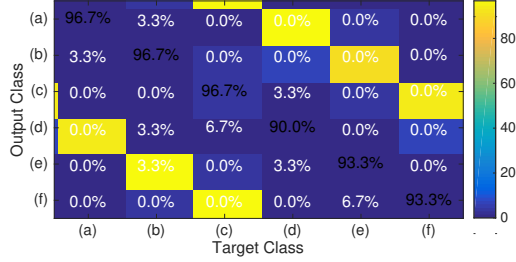
Fig. 15. Individualized confusion matrix for the six identified postures. The correlation between the postures (as shown in Figure 14), illustrates a $\approx 90\%$ classification accuracy. Similar postures incur misclassification due to changes in the patient's joint locations (such as the wrists or elbows).

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 applied across several patients with similar body shapes and sizes, obtaining reasonable results. The confusion matrix in Figure 16 shows the classification results of the postures provided by three individuals based on a pre-trained posture set formed from a single individual, with avg. accuracy ~90.62%.
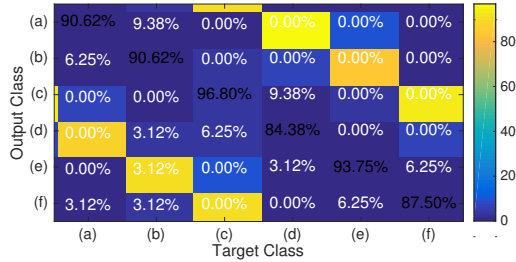

Fig. 16. Confusion matrix illustrating the accuracy of the posture estimation tested against a set of 3 patients that did not contribute to the training of the CNN used to perform the classification.

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 and illustrate the corresponding classification accuracies of the individualized experiment within Table I.

TABLE I
CNN POSTURE CLASSIFICATION PERFORMANCE

# of convolutional layers	1	2	3	4
Accuracy (%)	76.67	88.33	91.67	94.45
# of weights (millions)	1.2	2	2.8	3.11
Training time (minutes)	4.5	8.5	15	20.5

Evaluation and Discussion. There are three primary considerations employed within the design of these results that are addressed within this study: (1) the trained set is based on a discrete enumeration of skeletal postures, limiting the skeletal movement resolution, (2) the entire voxel volume is utilized without features, so large variance in body-type reduces accuracy, and (3) skeletal refinement algorithms 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, 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 potential future extensions of this method. The implemented volumetric reconstruction algorithm also provides a means of accurately modeling and visualizing the volumetric posture of a patient within an occluded region. This allows the this method to be applied to numerous additional medical imaging applications such as patient monitoring, gait analysis, and thermal distribution modeling for studies with similar occlusion constraints.

IX. 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 trained sleeping postures. 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, illustrated through the application to six predefined postures with an average classification accuracy of ~94.45% for an individual and an accuracy of ~90.62% for untrained patients.

REFERENCES

[1] N. Mohsin *et al.*, "Signal processing techniques for natural sleep posture estimation using depth data," in *2016 IEEE IEMCON*, Oct 2016, pp. 1–8.
[2] J. Shotton *et al.*, "Real-time Human Pose Recognition in Parts from Single Depth Images," *Commun. ACM*, pp. 116–124, 2013.
[3] M. Ye *et al.*, "Accurate 3D pose estimation from a single depth image," in *IEEE ICCV*, 2011, pp. 731–738.
[4] J. Shotton *et al.*, "Efficient Human Pose Estimation from Single Depth Images," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 2821–2840, 2013.
[5] S. Transue *et al.*, "Real-time tidal volume estimation using iso-surface reconstruction," in *ACM/IEEE CHASE 2016*, 2016, pp. 209–218.
[6] P. N. *et al.*, "Continuous and fine-grained breathing volume monitoring from afar using wireless signals," in *IEEE INFOCOM'16*, pp. 1–9.
[7] ——, "WiKiSpiro: Non-contact Respiration Volume Monitoring During Sleep," ser. ACM S3, 2016, pp. 27–29.
[8] ——, "Continuous and Fine-grained Respiration Volume Monitoring Using Continuous Wave Radar," in *ACM MobiCom*, 2015, pp. 266–268.
[9] F. Achilles *et al.*, "Patient MoCap: Human Pose Estimation Under Blanket Occlusion for Hospital Monitoring Applications," in *MICCAI*. Springer, Cham, 2016, pp. 491–499.
[10] X. Liu *et al.*, "Toward study of features associated with natural sleep posture using a depth sensor," in *IEEE CCECE*, 2016, pp. 1–6.
[11] S. Iwasawa *et al.*, "Real-time human posture estimation using monocular thermal images," in *IEEE FG*, 1998, pp. 492–497.
[12] I. Riaz *et al.*, "Human detection by using centrist features for thermal images," in *WSCG*. Citeseer, 2013.
[13] K. Yasuda *et al.*, "Thermo-key: human region segmentation from video," *IEEE Computer Graphics and Applications*, pp. 26–30, 2004.
[14] W. Wang *et al.*, "Improved human detection and classification in thermal images," in *IEEE ICIP*, 2010, pp. 2313–2316.
[15] A. Makadia *et al.*, "Fully Automatic Registration of 3d Point Clouds," in *IEEE CVPR*, 2006, pp. 1297–1304.
[16] A. Krizhevsky. *et al.*, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.
[17] N. D. Lane *et al.*, "DeepEar: Robust Smartphone Audio Sensing in Unconstrained Acoustic Environments Using Deep Learning," in *ACM Ubicomp*, 2015, pp. 283–294.
[18] N. Srivastava *et al.*, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, pp. 1929–1958, 2014.