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 with 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 and introduce a new occluded skeletal posture 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. This approach introduces disconnected skeletal definitions and allows correct posture estimation in highly ambiguous cases. We demonstrate the accuracy and utility of this approach by correctly identifying both standard sleeping postures and complex ambiguous skeletal configurations that currently not provided by existing methodologies.**

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 the application of these techniques are not well suited for applications that include visual obstructions such as sleep-based studies where patients are heavily occluded by both clothing and common forms of bedding including sheets and blankets. While recent depth-based imaging methods [5], [6] 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 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 fail to identify obscured skeletal joints. In this work we explore how the limitations of depth images can be supplemented through the integration of thermal imaging to introduce new methods in occluded skeletal posture estimation.

Modern digital imaging contains 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

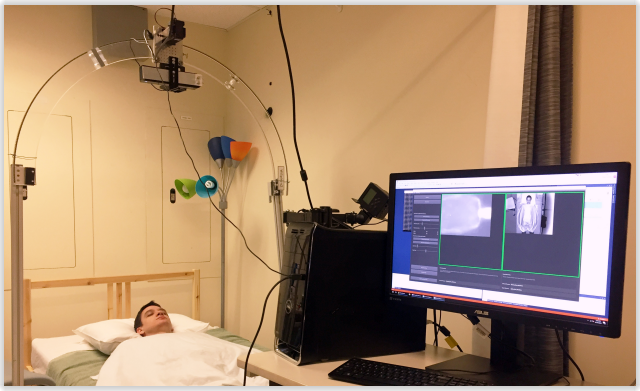

Fig. 1. Experimental setup for detecting occluded skeletal joints that defines a patient's posture with occlusions from standard bedding. Emphasis: The proposed thermal-based skeletal estimation prototype for both depth and thermal imaging devices to collect 3D point-clouds of the patient's occluded posture.

framework for both training and in medical practice these imaging techniques are not well suited 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 more reliable means of estimating occluded skeletal postures.

In this novel approach to occluded skeletal posture estimation we introduce three primary contributions: (1) we present 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 regions and (3) introduce a fine-grained skeletal posture estimation technique for identifying joint positions of visually obscured patients in sleep-based studies.

Based on these contributions, we provide a new method for generating occluded posture volumes through depth and thermal imaging, a correlation of these distributions with skeletal ground-truth postures, and a robust training method for identifying the skeletal postures of newly generated posture volumes. We then evaluate this new method by assessing its ability to correctly identify several common sleep postures and generate accurate skeletal joint positions based on a patient's occluded thermal distribution for instances where joint positions are highly ambiguous due to visual occlusions.

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 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 rapidly identify depth pixel information and their contribution to known skeletal joints and hand gestures introduced with the Kienct2. Modern skeletal estimation techniques are built around a similar premise and utilize an extensive number of newer devices that provide high-resolution depth images at 30[fps] or greater. 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 image in Figure 2 provides an illustration of the most common skeletal configurations and associated estimation results from recent techniques.

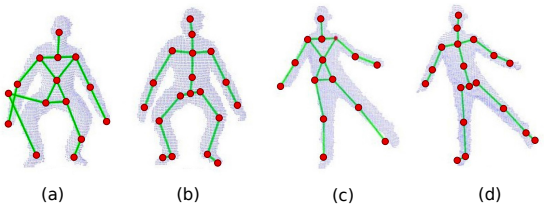

Fig. 2. Skeletal posture estimations from recent techniques from Kinect, Primesense OpenNI (a, c), and improvements (b, d) by [3] and that utilize depth-imaging to accurately identify joint positions in non-occluded applications.

These techniques have become increasingly robust and now provide highly accurate joint estimations within the boundary 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 remain invalid.

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 [7] 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 [8] and tracking generalized human behaviors [8], [9], [10] which include movement and very generic postures such as walking, lying, 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.

Occluded Skeletal Posture Estimation. Recent vision-based techniques have introduced alternatives that rely on a surface prior that provides a skeletal posture estimation that is 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. However, there are several potential problems with this approach: (1) the simulated cloth under gravity model may not provide realistic behavior due to the simplified configuration, (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 such as tucking or stretching.

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. 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. This is simply because the surface can be modified by simply moving a blanket while the patient remains stationary. This introduces ambiguous surface information that may incorrectly correlate the surface information with an incorrect posture.

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 both 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 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 intangible when an occluding medium is present.

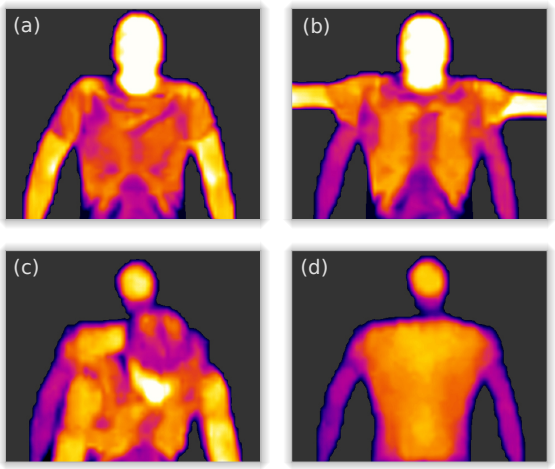

Fig. 3. Skeletal posture estimation challenges associated with thermal imaging. The image in (a) illustrates an ideal non-occluded thermal image but illustrates non-uniform thermal distribution of a patient's thermal signature, (b) provides an illustration of the residual heat left by the patient's arm movements, (c) illustrates thermal ambiguities of the patient, 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. Therefore to accurately identify the shape or distance of an object behind an occluding source using a thermal image, this distance function, the shape of the emitting object, the emission intensity, and the physical characteristics of all materials must be precisely modeled.

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, due to the introduction of an occluding material, 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 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. This is due to the additional layers absorbing a portion of the thermal radiation from the emitting surface based on the thermal conductance of the occluding materials, leading to various temperature levels based on the number of layer occlusions.

Intractable Heat-to-Surface Modeling. The basic premise of trying to identify and generate 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 accurate spatial definition of the emission surface.

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, the different thermal intensities lead to ambiguities between the patient's skin and surrounding materials.

Movement and Residual Heat. As a challenge primarily associated with thermal imaging, thermal contact and residual heat play a critical role in the image analysis of patient postures. During movement events and for an environment dependent time duration after the movement, thermal intensities may indicate false positives in posture estimations.

Heat Pockets. Heat pockets are defined as isolated regions of residual heat that maintain a high thermal intensity due to the thermal insulation of bedding materials after the patient has moved. The propagation of heat from a patient's skin to surrounding materials introduces heat residuals upon movement that may become trapped within the thermally insulated volume. This presents a significant challenge for thermal imaging due to the timespan of these residuals as they slowly dissipate through surrounding materials. These form high intensity thermal regions that may form isolated disconnected components within the thermal image that make a unique skeletal posture ambiguous.

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 with loosely defined constraints. This will provide what we define as the *posture-volume* of the patient. This volume is defined as the continuous region under the occluding surface that contains both the patient and empty regions surrounding the patient that are visually obscured. 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.

To illustrate this volumetric model and an associated skeletal posture, the images in Figure 4 provide an overview of an ideal posture volume and how the correlated skeletal estimation provided by a ground-truth estimate can be defined in terms of the patients volume in the occluded region and thermal distribution.

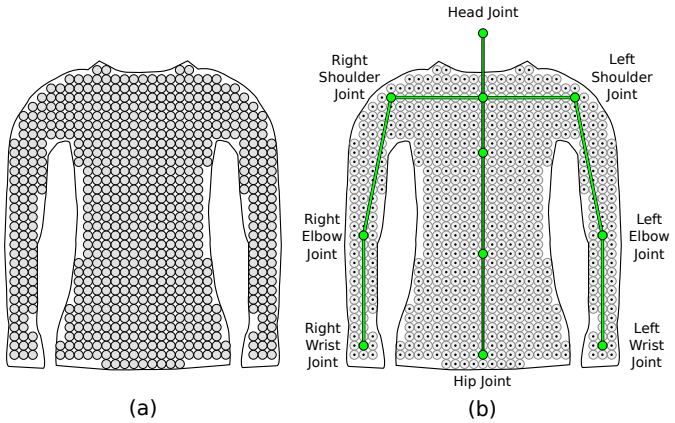

Fig. 4. Volumetric reconstruction of an ideal skeletal posture. The image in (a) illustrates the approximated volume of the patient's posture which may be drastically influenced by the occluding medium covering the patient. The image in (b) provides an illustration of the mapping between a voxel representation (black dots) of the volumetric data and the ground-truth skeletal estimate of the posture (illustrated as a set of bones and joints).

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 provides a complete 3D image of the patient's posture within the occluded region as an identifiable thermal distribution that can be assigned through training an associated skeletal estimates that may contain visually ambiguous joint positions.

Based on identification of this thermal-volumetric signature, our method provides a skeletal estimation that is based on the correlation provided between this new thermal distribution and the patient's skeletal posture introduced by our new ground-truth estimation. This ground-truth estimation is derived from modern motion capture techniques and provides a set of thermal markers that are used to identify the locations of the patient's joints within the thermal image.

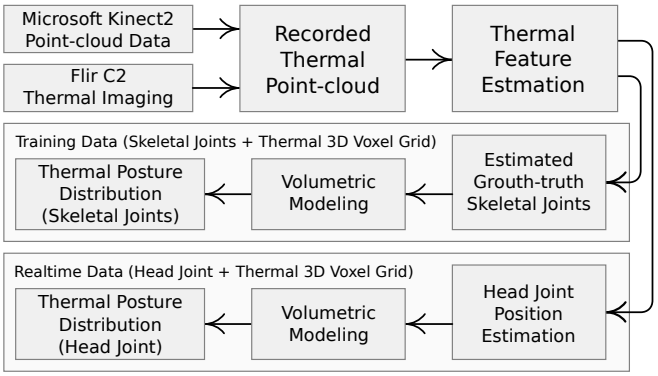

Fig. 5. Overview of the proposed approach for reconstructing the volumetric thermal data that contributes to the occluded skeletal posture estimation. This includes the generation of the volumetric data with the skeletal ground-truth for training and the real-time data with the provided head joint used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

B. Algorithm Overview

The premise of this approach is to reconstruct the volumetric thermal distribution of the patient which is unique for each potential posture and correlate this unique posture signature with an associated set of joints that define a close approximation of the patient's skeletal posture. This provides a robust method of identifying skeletal estimations on more reliable volumetric data that contains unique thermal patterns rather than relying purely on thermal features within the surface of the recorded point-cloud. The resulting correlation is then used as a training model to complete skeletal generation and detection.

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. This includes the Microsoft Kinect2 for depth imaging and the Flir C2 hand-held thermal imaging camera. The Kinect2 provides a depth-image with a resolution of 512x424 and the C2 contains an 80x60 thermal image sensor array which is upscaled to 320x240. To configure the 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.

A. Thermal Surface Point-clouds

The depth image generated by the Kinect2 provides a direct translation to a 3D point-cloud that contains all of the depth points in camera-space. Based on this point-cloud data, we integrate the thermal intensity at each point from the corresponding point within an up-sampled thermal image provided by the C2. 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 field of view possible.

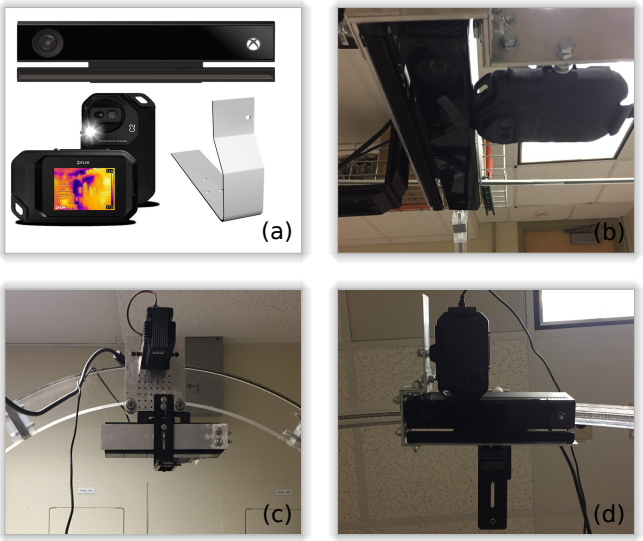

Fig. 6. Thermal posture device prototype. The two devices (Kinect2, C2) are mounted with a fixed alignment provided by the bracket shown in (a). The images in (b-d) illustrate the mount attached to the bed rail with both devices pointing to the surface of the bed. This minimalistic prototype provides the ability to collect precisely aligned depth and thermal images for skeletal posture estimation.

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 patients 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

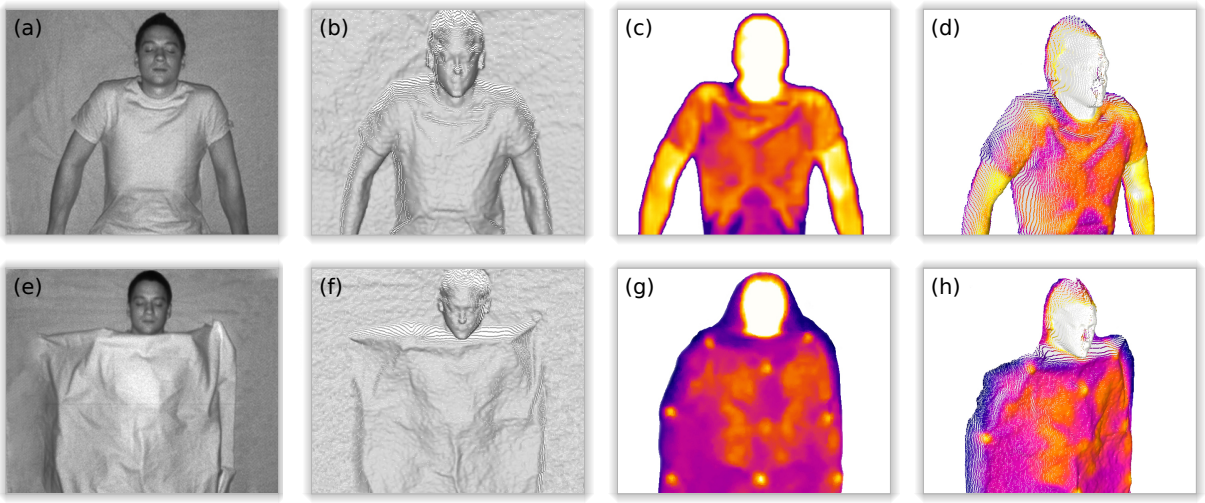

Fig. 7. Thermal surface point-cloud acquisition. The sequence of images illustrate the data collected from both the Microsoft Kinect2 and Flir C2 thermal devices to obtain thermal and surface point-cloud data. The images (a-d) illustrate the collection of the infrared, depth, thermal, and thermal surface 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. Using the proposed ground-truth estimation, we can assert known joint positions through occluding materials.

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 occluding surface is present. The image provided in Figure 8 illustrates the simple design of the training suit with the attached solid nickel spheres used in the training process.

Fig. 8. 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 occluded volume and joint positions (c).

The result of the thermal skeletal ground-truth is the product of a simple adaptive thresholding and connected-component algorithms that identified by the thermal intensity of the spheres. In the resulting thermal-cloud, the spheres appear as small white regions indicating the locations of the joint positions. 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 joint adjacencies, the system will automatically generate the required skeletal structure.For occluded joints, we introduce the notion of an incomplete skeletal structure as illustrated in Figure 9.

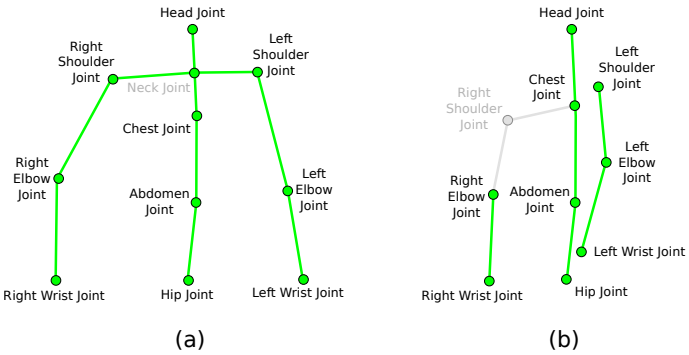

Fig. 9. Thermal skeleton ground-truth. The ground-truth skeleton presented in (a) illustrates a complete skeletal posture based on every supported joint being identified. The skeleton presented in (b) represents a different the patient in a left facing posture with their right shoulder joint occluded from the thermal image. This results in a disconnected skeletal posture.

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 within a limited region

at rest 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 10. This is achieved through the use of several assertions about the experimental setup: (1) the patient resides within the bounded region and is supported by a rest surface, (2) the occluding surface is supported by the patients body and does not penetrate through the volume of the body, (3) the human body is contiguous, and (4) the patient's face is visible.

A. Volume Enclosure

To begin the process of imposing constraints on the possible joint locations within the occluded region, we begin by enclosing the volume between the recorded depth image and the 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 the use of a depth image that defines the beds surface when the 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.

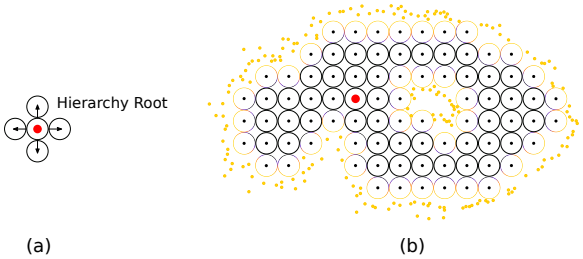

Fig. 11. Two-dimensional variant of the volumetric reconstruction algorithm. The image in (a) illustrates the hierarchy root and the propagation directions and (b) illustrates the limitation of the propagation by the surrounding point-cloud and associated thermal intensities.

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 additional addition 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 11 illustrates this thermal sphere-packing algorithm demonstrated on a two-dimensional point boundary with the staring position residing within the volume.

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 $TEGI(t, d)$:

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

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

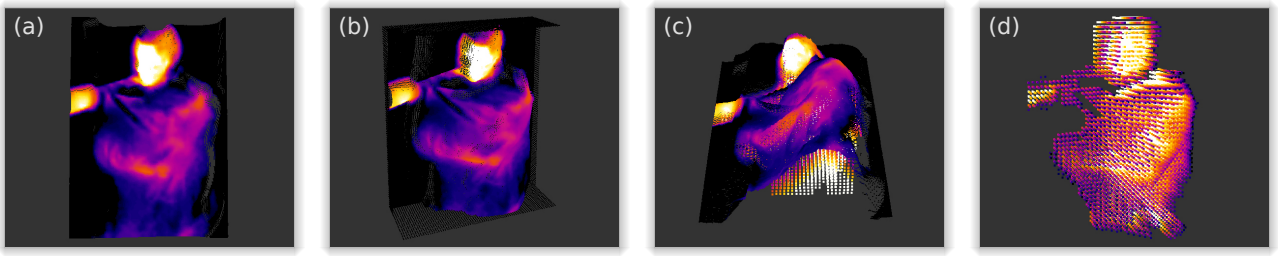

Fig. 10. Volumetric thermal model process overview. The image in (a) illustrates the raw thermal cloud, (b) illustrates the enclosed region of this cloud, (c) illustrates the generated internal thermal distribution of the patient, and (d) provides the result of both the reconstruction and the thermal propagation through the enclosed volume. The thermal distribution in (d) will then be provided to the training algorithm with an associated skeletal estimation.

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.

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 [11] 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 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.

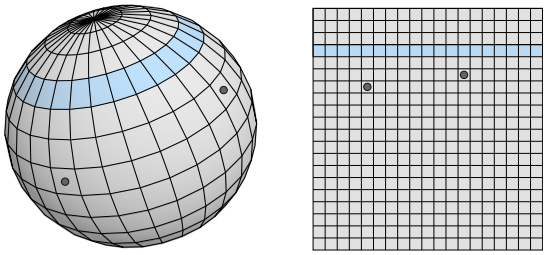

Fig. 12. Extended Gaussian Image (EGI) spherical mapping [11]. For each thermal point within the recorded thermal point-cloud, the projection of the point will produce a location on the unit sphere that will reside within a bounded surface region. These surface regions are defined by the height and width of the EGI map (b). The corresponding surface regions in (a) are displayed in the two-dimensional representation in (b).

The 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 $\mathcal{N}$:

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

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 a means for estimating how the internal heat is propagated to the surface of the depth image which can assist modeling boundaries of the patient's joint positions.

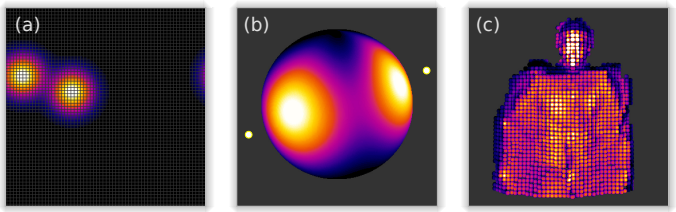

Fig. 13. Thermal Extended Gaussian Images for the distribution of heat due to surrounding thermal points. The image in (a) represents the discrete map of the sphere surface that contains the thermal contribution of two points. The image in (b) illustrates the TEGI in 3D space with the two contributing points.

The TEGIs are evaluated for each sphere in the spherical hierarchy that reside within the surface of the thermal cloud that contain points with thermal data in their local neighborhood. The resulting thermal intensity of each sphere is then used as the seed for propagating the observed heat through the patient's posture volume to provide an approximate linear thermal propagation model of the patient.

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

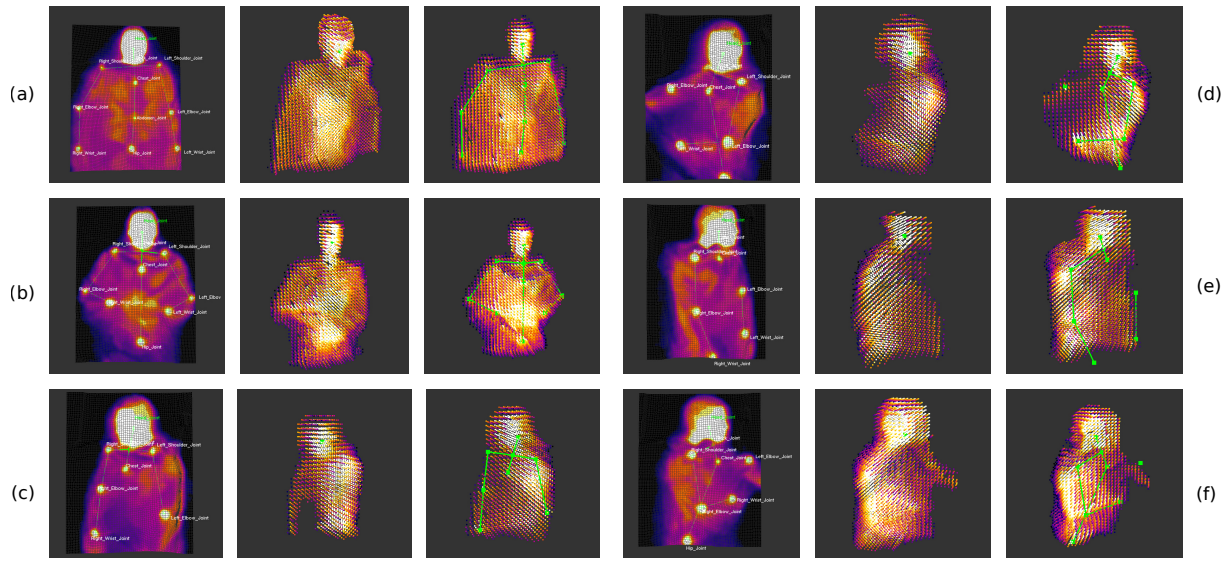

Fig. 15. The results of skeletal posture estimation. **This is a place holder for now. We need to gather more data.**

distribution of the internal volume of the patient used to represent the thermal distribution of a unique posture.

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.

A. Convolutional Neural Network Training

There are several types of training methodologies and learning models that have been designed for three-dimensional medical image classification. Of these methods, Convolutional Neural Network (CNNs) [12] and Deep Neural Networks (DNNs) [13] are most commonly used methods for identifying complex structures within 3D images.

Neural Network Structure. 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. This is due to our models dense representation rather than feature-based estimations which are better suited for DNN methods. Therefore we allow the CNN to generate features through learnable 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 [14].

Convolutional Layers. The key building blocks of the convolutional neural networks is the use of convolutional layers including sets of learnable filters. For each convolution layer, the size of the convolutional kernel decides the shape and number of feature maps used in convolution operation. There is no analytical method to determine the optimal number of convolutional layers for a given application, therefore our network structure is determined empirically.

Training Model. We trained CNNs to detect 6 postures of human based on our generated thermal voxel grid images. The label (one of 6 postures) is assigned for each image. 60 thermal voxel grid images are used for training while 180 other images are used for testing. Note that overfitting is one of the key concern of this training process. The overfitting is avoided through 2 main steps: First we apply the *Dropout* algorithm to randomly drop units (along with their connections) from the neural network during training [15]. This prevents neurons from co-adapting. Second, cross-correlation is applied to stop the training when the cross-validation error starts to increase. 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 [12], [16].

VIII. Experimental Results

In this section we present an overview of the result obtained by using the proposed thermal distribution identification algorithm used to assign an observed heat distribution to a known skeletal estimate. Using the proposed method we can obtain the skeletal estimate of the patient that defines the *closest* skeletal match possible based on the set of trained thermal distributions.

Standard Posture Estimation. This section describes the results obtained by estimating the human posture using CNN training for standard postures with skeletal joint estimates.

The input data includes 4 properties: [Train Images|Train Labels|Test Images|Test Labels]. In that, the |Train Image| contains a struct of multiple objects, each object is represented by a 3D image defined by the three-dimension voxel data and the |Train Labels| includes the corresponding posture labels. We collect 40 data set for each postures. There are 240 data set in total, we use 60 data set for training (10 data for each posture) and 180 data set for testing with different variations generated by participants. The confusion matrix illustrating in Figure 16 shows the performance of our posture detection algorithm. As can be seen from the results, the postures are estimated correctly with up to 96.7% of accuracy.

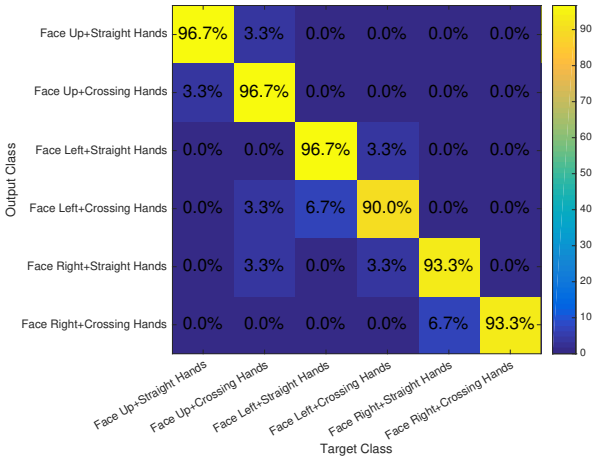

Fig. 16. The confusion matrix of estimating human postures.

Impact of Training Network Structure. The introduction of additional layers within the CNN improves the performance of detection significantly, but we still observe diminishing returns. We tested the CNN from 1 to 4 convolutional layers and illustrate the results within Table I. While we obtain high accuracy of posture detection, the proposed training algorithm requires much more computational time when we introduce 3 and 4 fully connected layers for training. However, for real-time applications this method can be used due to the minimal execution time required to perform a classification on the trained neural network.

TABLE I
COMPARISON OF PERFORMANCE ON POSTURE DETECTION FOR A CNN WITH DIFFERENT NUMBER OF CONVOLUTIONAL LAYERS

# of convolutional layers	**1**	**2**	**3**	**4**
Accuracy (%)	76.67	88.33	91.67	94.44
# of weights (millions)	1.2	2	2.8	3.11
Training time (minutes)	4.5	8.5	15	20.5

Medical Significance. The implemented skeletal estimation algorithm also provides a means of accurately modeling the volumetric posture of a patient within an occluded surface as illustrated in Figure 15. This allows the proposed method to be applied to numerous additional medical imaging applications. These include long-term sleep studies, gait analysis, human sleep behavior analysis, sleep staging, and thermal distribution modeling of the body.

IX. Conclusion and Future Work

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 accurate skeletal joint estimates for applications that include surface occlusions. By extending this approach to define a patient's unique thermal distribution within an occluded region, we have also introduced a new means for accurately visualizing a patient's posture within an occluded region. Based on the provided training set used to estimate six standard sleeping postures with the patient's skeletal estimation, this technique can be extended to incorporate larger training sets to provide estimates for additional patients without additional training. This represents a large portion of the intended future work that will utilize this method for incorporating several additional participants within the training process. In its current state this work provides a first step towards using thermal imaging to provide accurate skeletal estimates during real-time posture analysis without disturbing the patient.

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 *International Conference on Computer Vision*, Nov. 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] F. Achilles *et al.*, "Patient MoCap: Human Pose Estimation Under Blanket Occlusion for Hospital Monitoring Applications," in *MICCAI 2016*. Springer, Cham, 2016, pp. 491–499.

[6] X. Liu *et al.*, "Toward study of features associated with natural sleep posture using a depth sensor," in *IEEE CCECE*, 2016, pp. 1–6.

[7] S. Iwasawa *et al.*, "Real-time human posture estimation using monocular thermal images," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 492–497.

[8] I. Riaz *et al.*, "Human detection by using centrist features for thermal images," in *International Conference Computer Graphics, Visualization, Computer Vision and Image*. Citeseer, 2013.

[9] K. Yasuda *et al.*, "Thermo-key: human region segmentation from video," *IEEE Computer Graphics and Applications*, pp. 26–30, 2004.

[10] W. Wang, J. Zhang, and C. Shen, "Improved human detection and classification in thermal images," in *2010 IEEE International Conference on Image Processing*, Sep. 2010, pp. 2313–2316.

[11] A. Makadia *et al.*, "Fully Automatic Registration of 3d Point Clouds," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, pp. 1297–1304.

[12] A. Krizhevsky *et al.*, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012.

[13] ——, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097–1105.

[14] N. D. Lane *et al.*, "DeepEar: Robust Smartphone Audio Sensing in Unconstrained Acoustic Environments Using Deep Learning," in *ACM Ubicomp*, 2015, pp. 283–294.

[15] N. Srivastava *et al.*, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, pp. 1929–1958, 2014.

[16] O. Abdel-Hamid *et al.*, "Exploring convolutional neural network structures and optimization techniques for speech recognition," in *Interspeech 2013*. ISCA, 2013.