Real-time Tidal Volume Estimation using Iso-surface Reconstruction

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—Breathing volume measurement has long been an important physiological indication widely used for the diagnosis and treatment of pulmonary diseases. However, most of existing breathing volume monitoring techniques require either physical contact with the patient or are prohibitively expensive. In this paper we present an automated and inexpensive non-contact, vision-based method for monitoring an individual's tidal volume, which is extracted from a three-dimensional (3D) chest surface reconstruction from a single depth camera. In particular, formulating the respiration monitoring process as a 3D space-time volumetric representation, we introduce a real-time surface reconstruction algorithm to generate omni-direction deformation states of a patient's chest while breathing, which reflects the change in tidal volume over time. These deformation states are then used to estimate breathing volume through a per-patient correlation metric acquired through a Bayesian-network learning process. Through prototyping and implementation, our results indicate that we have achieved 92.2% to 94.19% accuracy in the tidal volume estimations through the experimentation based on the proposed vision-based method.

I. INTRODUCTION

Monitoring patient tidal volume over time can be effectively used to extract important indicators of pulmonary medical conditions that are identifiable through a patients respiration patterns. The applicability, portability, and accuracy of these techniques based on this premise have been extensively surveyed [1]. The inherent objective of these approaches is to accurately monitor respiration rate or tidal volume to identify respiratory disorders and eradicable conditions such as pulmonary edema, pneumothorax, and chronic diseases such as tuberculosis, cystic fibrosis, and chronic pulmonary disease (COPD). Several approaches introducing a vast array of techniques have been proposed over the course of several years utilizing different sensor types including: camera-based [2], [3], Doppler radar-based [4], laser-based measurement [5], infrared (IR) imaging [6], magnetometers [7], and high-resolution accelerometers [8], [9].

The ability to provide self-contained systems for the detection of these conditions plays a critical role for at-home care patients and those without access to typical in-clinic solutions for regular monitoring, a domain where intuitive vision-based techniques provide an unobtrusive cost-effective solution. Several contributions within the vision-based respiration modeling domain have been aimed at efficiently extracting a patient's breathing rate by identifying and measuring visible chest displacements over time. Recent developments in affordable
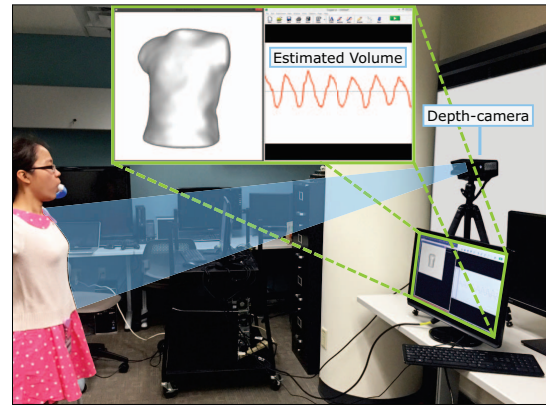
Fig. 1. Proposed system environment for real-time surface-based tidal volume monitoring. Emphasis: screen illustrating the real-time surface reconstruction and estimated tidal volume during the patient monitor training process. This setup illustrates the non-invasive methodology proposed by our vision-based tidal volume estimation technique.

high-resolution depth-imaging hardware have facilitated developments using infrared projection [10] and depth-imaging techniques [11] for monitoring patient respiration rates in clinical and sleep-based studies. These contributions illustrate the utility in providing unobtrusive vision-based approaches facilitating respiratory pattern estimation for patient care; however they contain manual configuration procedures for chest region identification and are limited in their accuracy due to low resolution 2D depth-image orthogonal displacements.

While vision-based monitoring systems also suffer from inherent challenges related to occlusion, monitoring distance, and clothing related interference, this technique remains one of the least intrusive (non-contact) methodologies to monitoring a patients respiration patterns in real-time. However, the requirements that occur in any vision-based monitoring technique must be counterbalanced by the quality and accuracy of the estimation model. The premise of any vision-based approach that provides highly accurate estimates must rely on the assumptions that the patient resides within an unobstructed environment, that their clothing is classified as *form-fitting* where we define form-fitting as any light material that loosely adheres to the patients underlying chest surface, and the patient resides within the device field-of-view (FOV).

In this work we propose a methodology that facilitates unobtrusive monitoring of respiration patterns of a patient observed through an individual depth-imaging device with skeletal-tracking [12], [13] data as a self-contained automated respiratory monitoring framework shown in Figure 1. This method

presents a novel approach for direct tidal volume estimation derived through 3D iso-surface reconstruction bridging the gap between 2D depth-based monitoring and three-dimensional space-time modeling introduced by prior approaches [2]. This approach extends prior non-contact vision-based techniques that generate three-dimensional chest models [14] to provide a new means of extracting breathing characteristics specific to the monitored individual's chest deformations represented by *water-tight* surface reconstructions. These behaviors and reconstructions are then used to accurately model the patient's unique respiratory patterns and directly measure the patient's corresponding respiration rate and tidal volume.

To increase the accuracy of the estimated tidal volume based on the direct measurement of an individual's characteristic respiratory patterns, we introduce a regenerative Bayesian-based neural network to improve the correlation between the calculated volume and the patient's unique breathing characteristics. This process provides a per-patient correlation between the measured tidal volume and the patient's actual tidal volume recorded with a spirometer. This technique is consolidated into our automated monitoring process as an initial training period within the real-time monitoring procedure to provide accurate respiratory rate monitoring and tidal volume estimation.

II. RELATED WORK

Numerous techniques have recently been proposed for monitoring breathing, they are not only supported for clinical purposes but also for in-home well-being monitoring. Polysomnography (PSG) is a gold standard technique to breathing monitoring, but it is complicated, expensive and is not available for testing in most rural areas [15]. Using other traditional techniques such as spirometer and nasal thermocouples are obtrusive and not practical for continuous monitoring. Prior straightforward approaches place accelerometers and gyroscope sensors on the patient's chest and measure the chest movement according to the acceleration change. These acceleration changes are then used to infer the breathing volume [8]. However, this technique is also obtrusive and less accurate due to the effects from drifting noises of acceleration data itself. Wireless-based techniques has also attracted attention for breathing monitoring purposes [4]. However, most of wireless-based techniques are used to estimate the breathing rate, which cannot provide fine-grained information of breathing behavior and complex chest deformations. Breathing volume is also approximated by observing the effect of frontal chest movements on the phase of the reflected-off wireless signal when a radar beams a continuous wave signal directly to the patient's chest [16]. However, as different areas within the chest move differently, these techniques are not suited for capturing complex surface deformations of the patient's chest as they breathe.

Vision-based respiratory monitoring is a relatively new prospect with respect to real-time fine-grain tidal volume estimation. Recent depth-imaging hardware developments (*e.g. Kinect-2*) have provided vision-based techniques with the high resolution depth images that are required to adequately capture

the fine movements of a patient's chest during the respiration process. Recent developments in depth-image processing have provided the basis for recording the periodic displacement of a patient's chest wall or surface from a static reference plane [3]. The summation of this frontal displacement with the depth-image can then be used to estimate the respiration rate and tidal volume of a patient [2], citing a high correlation between the recorded volume and the corresponding spirometer measurement. Similarly, techniques proposing a method of extracting the tidal volume of a patient by reconstructing an orthogonal representation of a triangulated surface defined by the depth-cloud collected by the Kinect-1 [17]. In this work the process of identifying the thoracoabdominal domain has been automated by utilizing the skeletal information provided by the device; however this only identifies a limited region of the chest surface from which the tidal volume is inferred.

The fundamental deformation model presented within each of these techniques is based on an orthogonal projection of the chest displacement in the direction of the monitoring device using a time-series alteration of the three-dimensional shape. While this model allows for some flexibility in the positioning of the patient, it does not accurately represent the natural surface deformation imposed on the patient's chest. In this work we present an alternative model that utilizes an omni-directional expansion to more accurately represent monitored chest deformations.

III. METHOD OVERVIEW

The accurate estimation of a patient's tidal volume using a vision-based technique is dependent upon both the model of respiratory deformation patterns and the correspondence relation used to provide a correlative link between this behavior and the actual tidal volume. The challenges presented in obtaining an accurate estimation result are derived from the correlation of the models from the true deformation behavior and the means of accurately obtaining the prerequisite correspondence for populating the models estimation basis. In the respiratory monitoring framework we propose, we address these challenges by introducing a two phase correspondence model from which the chest surface deformations, respiration rate, and tidal volume can be effectively extracted and estimated on a per individual basis. This estimation is initially obtained using direct 3D volume measurement and then improved using a per-patient trained correlation function.
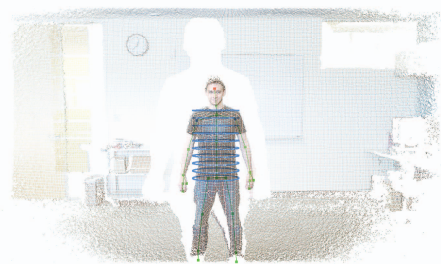

Fig. 2. Color point-cloud acquired from the device with the both the skeletal and clipping cylinder super-imposed. Any vertical posture within the devices field-of-view (FOV) is valid with our approach.

To obtain these individualized respiratory characteristics, we have developed a methodology for extracting a complete volumetric iso-surface that includes the deformation behavior of the patient's left thorax, right thorax, and abdominal region. We also introduce a new deformation model that provides a closer representation of a naturally expanding chest cavity to increase the accuracy of a patient's estimated tidal volume. This respiration model is then combined with a adaptive correspondence model that utilizes a Bayesian-based neural network to populate a regenerative tidal volume estimation.

A. Omni-directional Deformation Model

The proposed respiratory model is fundamentally composed of the accurate reconstruction of a volumetric region enclosed by an iso-surface that describes both the deformation characteristics of a patient's chest and the change in volume of the patient's chest. The premise of our omni-directional model is based on the accurate approximation of a solid volume by its *characteristic function* formed from a set of unordered, oriented points that allows us to extract the iso-surface that describes these characteristics.

In this approach, we minimize the mobility of the patient during the monitoring process to employ this omni-directional chest deformation model to form a more accurate basis for the correlation between a patient's chest deformations and the corresponding tidal volume. This also allows us to consider the chest deformations specific to the monitored patient within our estimations providing a better model to infer the associated tidal volume. In this section, we describe the basis of our model as compared with prior techniques and provide an derivation of how this model is applied to form a more accurate representation of the chest deformations observed during a patient's breathing cycle.

Prior techniques for modeling chest movement utilize orthogonal deformation models of a patient's chest surface to infer the correlation between the monitored chest movements and the corresponding tidal volume. These models are based on the orthogonal movement of the chest within a depth-image as displacements. The change in these displacements is then utilized to form a correlative relation between the chest displacement and the estimation of the patient's tidal volume. Our method is motivated by the observation that this deformation model does not accurately represent the known physiological displacements of a human lung during the respiration process. The images in Figure 3 illustrate the difference between an orthogonal displacement model and the proposed omni-directional model.

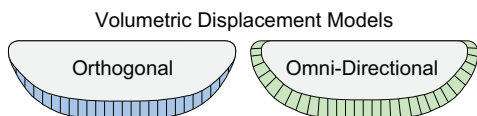#### Volumetric Displacement Models

Fig. 3. Comparison (top sectional view) of existing chest displacement models and the proposed omni-directional deformation model. An omni-directional model (right) provides a closer approximation of the natural chest displacements within the patient's chest during the respiration process when compared to an orthogonal model (left).

An omni-directional deformation pattern provides a closer approximation of the true displacements imposed on a patient's chest surface as they breathe. This is formulated based on the observation that the displacement incurred while breathing effects the estimated tidal volume which is a function of the expansion of the left and right thorax (*e.g. the chest is modeled as balloons rather than a set of uniform displacements*). Using this observation, we aim to increase the accuracy of the deformation model that is used to derive the correspondence between chest deformations and the estimated tidal volume.

The derivation of our model based on the established methodology of reconstructing solid model surfaces from unordered, orientated, point sets [18]. We then illustrate the application of this method as a means to accurately estimating a patient's tidal volume based on the volumetric changes in the patient's chest model. In this formulation we denote the patient's chest $C(t)$ as a three-dimensional solid with volume $V(t)$ contained within the closed boundary surface $S(t) \subset \mathbb{R}^3$. This is derived from the fact that the *Divergence Theorem*:

$$\iiint_{V(t)} \nabla \cdot \vec{F} \, dV = \oiint_{S(t)} \langle \vec{F}, \vec{n} \rangle \, dS \qquad (1)$$

allows the volume integral of the solid chest region to be expressed as the surface integral which can be approximated using Monte-Carlo integration assuming discrete uniform surface sampling where $\vec{F} = (F_x, F_y, F_z) : \mathbb{R}^3 \to \mathbb{R}^3$ and $\vec{n}_i$ is the estimated surface normal at point $\vec{p}_i$:

$$\iiint_{V(t)} \nabla \cdot \vec{F} \, dV \approx \frac{|C|}{N} \sum_{i=1}^{N} \langle \vec{F}(\vec{p}_i), \vec{n}_i \rangle \qquad (2)$$

The aim of this technique is to reinterpret the characteristic function of this solid region as a set of volumetric integrals that can be computed as a summation over a set of surface samples. The characteristic function of the patient's chest region, denoted as $\chi_c(t)$ is a function that defines the solid volume $C(t) \subset \mathbb{R}$ by providing a function that evaluates to one within the boundary $S(t)$ and zero otherwise. The discrete form of the characteristic equation expressed in terms of Fourier coefficients can be defined as:

$$\hat{\chi}_c(l, m, n) = \int_{p \in S(t)} e^{-i(lp_x + mp_y + np_z)} \, dp \qquad (3)$$

Using the proposed application of the Divergence Theorem, it can be shown that due to expressing the Fourier coefficients as volume integrals, the evaluation of the Fourier coefficients of the characteristic function can be computed using the Monte-Carlo approximation:

$$\hat{\chi}_c(l, m, n) = \frac{1}{N} \sum_{i=1}^{N} \langle \vec{F}_{l,m,n}(\vec{p}_i), \vec{n}_j \rangle \qquad (4)$$

such that the vector valued function: $\vec{F}_{l,m,n} : \mathbb{R}^3 \to \mathbb{R}^3$ adheres to the condition: $(\nabla \cdot \vec{F}_{l,m,n})(x, y, z) = e^{-i(lx+my+nz)}$

and the the functions $\vec{F}_{l,m,n}$ whose divergences are equal to the complex exponentials. The inverse Fourier Transform of these coefficients is then computed through a convolution of the oriented samples through a voxel grid to extract the solids characteristic function [18].

From the accurate reconstruction of the surface $S(t)$ that bounds this solid region through the Monte-Carlo approximation, the basis of our omni-directional model provides a high resolution approximation of the deformations observed during the respiration process. Based on this approach, we aim to provide a more accurate estimation of the patient's tidal volume due to the more accurate representation of the patient's chest deformations.

B. Chest Volume Extraction

Non-contact based methodologies inherently require a means of identifying the patient's position and orientation in space as a prerequisite to estimating the tidal volume that corresponds to the observed chest movements. Automating this process provides consistency in the region of interest monitored for surface changes and limits additional requirements imposed on the patient during the monitoring process. The automation of this process also eliminates the requirement of strictly limiting the patient's position to a pre-configured region of interest. Rather we build on the premise that the skeletal data can be utilized for automating the process of identifying the patient's chest region [17] and exploit this information to simplify the monitoring process.

In this section we describe the process of identifying and extracting the patient's chest region to calculate the volume of the deformable surface that describes the respiration patterns of the patient. The image in Figure 4 illustrates this process.
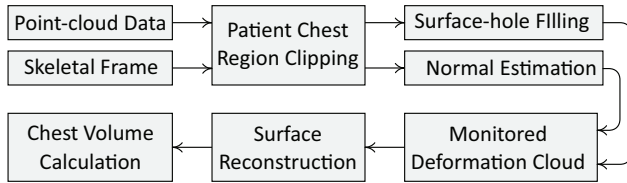

Fig. 4. Overview of the proposed approach to reconstructing the patient's chest surface in real-time. Each of the identified steps must be recalculated for each frame during the monitoring process. This provides an active representation of the patient as they are monitored and the resulting surface deformations closely illustrate the patient's breathing state.

The basic premise for reliably detecting the chest surface of the patient is derived from the acquisition of the *sampled* depth-image $D_s(t)$ (depth samples per-timestep) containing the patient and the raw skeletal data. Based on the forward orientation of the patient, assuming no occlusions, we consider the skeletal information as a basis for interpreting a chest subset c, denoted as $D_c(t)$, of the n-sampled depth image $D_s(t)$ as the chest region c of the patient at time t.

The subset of pixels $p_{ij} \in D_c(t)$ that reside within the cylindrically clipping region c, defined by the patient's skeletal structure and position, contribute to the definition of the patient's chest region. From this point, the objective is to form a representation of the patient's entire chest region as an enclosed volume defined through a point-cloud containing oriented points that approximate the patient's chest deformation states as a function of time, referring to this surface approximation as the volumetric deformation-cloud $\mathcal{P}(t)$. The samples collected from the depth-image, converted into three-dimensional coordinates, lack orientation vectors that approximate the curvature of the patient's chest. Therefore, in our reconstruction process we must generate accurate estimates of these normal vectors. Since timing is critical to the viability of this technique, we have proposed a reliable algorithm based on standard stencil techniques to approximate the surface normals within an organized point-cloud, covered in Section V.

Since the depth-image chest region subset contains a significant portion of occluded surfaces, including clipped regions and the patient's back, we propose a simple methodology for filling these occluded regions. This process is required to generate an enclosed volume to reconstruct a volumetric mesh of the patient's chest. In the process of approximating the surface of the occluded regions within the volumetric deformation-cloud, we employ a planar projection technique for mapping a patient's chest points as a fixed back surface and introduce a convex-hull based projection algorithm for filling the remaining clip holes (*e.g. neck, waist, etc.*). The aggregation of the chest, back, and generated clip-region points form the state of the volumetric deformation-cloud that is then used as the input to the iso-surface extraction algorithm. The overview of the proposed method is presented in Algorithm 1, where $B(t)$, $N(t)$, $W(t)$ represent the set of back, neck, and waist points respectively, $\mathcal{P}(t)$ is the volumetric point cloud, and $\mathcal{S}(t)$ is the reconstructed chest surface mesh.

Algorithm 1: Chest Mesh Volume Extraction (each time-step)

input : $D_s(t)$ - n-Sampled Depth-image
$\quad\quad\quad$ $S(t)$ - Patient Skeletal State
output: $\mathcal{V}_m(t)$ - Iso-surface Mesh Volume

foreach $p_{ij} \in D_s(t)$ **do**
$\quad$ **if** $p_{ij} \in c$ **then**
$\quad\quad$ $C(t) \leftarrow p_{ij}$
$\quad$ **end**

$B(t) \leftarrow \text{PlanarProjection}(\hat{b}, \alpha, C)$
$N(t) \leftarrow \text{ConvexPlanarProjection}(neck\_joint, E_c(t))$
$W(t) \leftarrow \text{ConvexPlanarProjection}(waist\_joint, E_c(t))$

$\mathcal{P}(t) \leftarrow \cup\ C(t)\ \cup\ B(t)\ \cup\ N(t)\ \cup W(t)$
$\mathcal{S}(t) \leftarrow \text{Iso-SurfaceExtraction}(\mathcal{P}(t))$
$\mathcal{V}_m(t) \leftarrow \text{SignedTetrahedralVolume}(\mathcal{S}(t))$

return $\mathcal{V}_m(t)$

The generation of this deformation model over time describes the deformation characteristics of the patient's chest that provides a correlation to the associated tidal volume. From the voxel-based surface reconstruction process, the generated triangulated mesh that represents the patient's chest volume $\mathcal{V}_m(t)$, is directly calculated using the signed tetrahedral volume algorithm [19].

Since the resulting surface reconstruction contains a sig-

nificant volume (including tissue and bone), we denote the volume initially recorded during the monitoring process as the base volume V_0. This value will then be subtracted off of all subsequent volume calculations to provide the discrete value dV for each time-step. Since this represents the form of our deformation correlation to tidal volume, dV is equivalent to the patient's tidal volume. In section IX-A, we extend this method through training to achieve a more accurate estimation.

IV. CHEST-SURFACE ACQUISITION

The acquisition of a depth-image from any infrared monitoring device incurs a natural variance in the depth measurements that are obtained within a single frame. In the instance of depth-imaging devices, the depth error associated with each pixel p_{ij}, is a function of the distance to the reflective surface being monitored as well as the surfaces material properties. Additionally, each pixel must be classified as part of the patient or as part of the background. The natural fluctuations within this process and depth measurement errors can degrade the accuracy of our tidal estimation. Therefore, in this section we cover the implementation of the cylindrical clipping region (Figure 5) and the associated pixel history tracking algorithm provided to minimize high-frequency pixel fluctuations.

The clipping cylinder that identifies the patient's chest region is defined through an automated process based on the subsection of a conventional skeletal frame illustrated in Figure 5.a. Specifically, the base of the cylinder is positioned at the hip joint h, and extends to the neck joint n. The radius of this cylinder is defined by the average distance of both the left l and right r shoulder joints. The generalized construction of this cylindrical clipping volume provides a viable heuristic for identifying the patient's chest volume bound by the accuracy of the skeletal joint estimations.
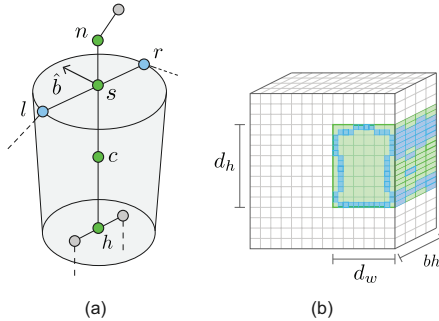

Fig. 5. Clipped skeletal structure provided by the Kinect-2 with our associated clipping cylinder (a). The depth-image bit history within the clipped region is utilized for removing depth measurement fluctuations belonging to the patient's chest surface (b).

To alleviate the natural fluctuation of the depth-image pixels that are determined to be part of the patient's body, but reside within the edges of the clipping region, we provide a simple stability scheme based on pixel tracking history. A visualization of this pixel-history is provided in Figure 5.b. If the tracking history of the pixel p_{ij} is saturated (continuously tracked) for the entire bit history length (bh), then it will contribute to the definition of the generated deformation-cloud.

This reduces the impact of fluctuating pixels as they are automatically culled from the background samples.

V. STENCIL-BASED NORMAL ESTIMATION

The process of reconstructing the surface of the scanned chest region requires every sample collected within the depth-cloud to have an estimated orientation that approximates the curvature of the surface. Normal estimation for surface reconstruction is a well studied research topic and several normal estimation techniques [20] have been developed and successfully employed within the Point Cloud Library (PCL) with widespread use. These techniques include k-neighbor or radial search for estimating normals for unordered point clouds [21]. and integral image normal estimation based on ordered depth-images [22]. However, in the process of clipping the patients chest region from the acquired depth-cloud, these techniques are ill suited for two reasons: (1) the additional computational cost associated with an unordered point set occupies unnecessary frame-time when adjacency information is known and (2) integral image techniques provides highly consistent estimated normals for ordered point-sets, but is not applicable in the instance of calculating reliable normals for the points that compose the edge of the clipped chest region due to the border of its rectangular region.

Due to the limiting factors of these recently developed normal estimation techniques in their application to the depth-cloud within the patient's clipped chest region, we employ an iterative stencil-based technique to accurately estimate all surface normals, including edge points and corners with missing adjacent neighbors. This is accomplished using the standard technique of sampling neighbor points to obtain an averaged cross-product that estimates $\hat{n}_{ij}$ at point p_{ij}.
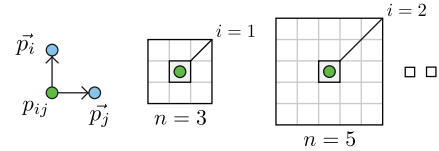

Fig. 6. Stencil-based normal estimation for different stencil sizes $n = 3, 5, ...$ to estimate the normal $\hat{n_{ij}}$ at point p_{ij}. As the stencil size is increased, the number of samples per point P_{ij} increases to contribute the surrounding area to the normals orientation.

The implemented technique relies on a stencil-based neighbourhood selection algorithm that calculates the cross-product of the current point p_{ij} with its surrounding neighbours. Based on the size of the stencil, c concentric squares are formed around the point where: $c \geq 3$ and $c \% 2 = 1$.

$$\hat{n}_{ij} = \sum_{i=1}^{c} \frac{\vec{p_i} \times \vec{p_j}}{|\vec{p_i} \times \vec{p_j}|} \qquad (5)$$

This calculation is then repeated through iterative clockwise rotations to provide an averaged normal estimation. For each concentric square at level i, each of the possible cross-products are calculated. Edge cases are handled by the 2D generation of valid point indices within the depth-image. After i iterations, the current sum is normalized to obtain the estimated normal of the surface at (i, j). The total number of cross products

performed for the given stencil size n is provided by: $\sum_{i=3}^{n} 2^i$ with a normalization. An illustration of this algorithm for $n = 5$ is illustrated in Figure 7.
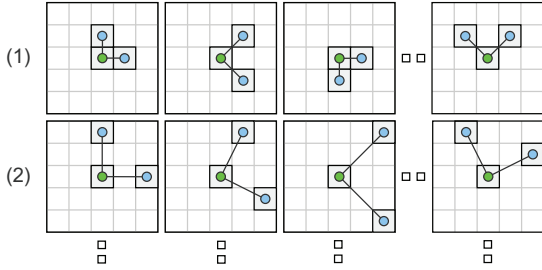

Fig. 7. Stencil-based normal estimation with a stencil size: $n = 5$ and the corresponding clock-wise crossproduct sampling indices for the central point at p_{ij}. Each row (i) illustrates iteration i of the algorithm for evaluating the cross-products at level i. All sampled cross-products from all levels are summed and then normalized to derive the estimated surface normal $\hat{n}_{ij}$.

A consideration for this normal estimation technique is addressed by the use of the integral image technique for calculating smoothed normals without relying on large stencil sizes. Naturally, larger stencil sizes require a higher number of samples, thus degrading the execution time of this technique. However, when using small stencil sizes (*e.g.* 3, 5), the algorithm provides normals that approximate the results provided by the PCL unorganized point set algorithm while generating accurate edge normals. An illustration of the normals estimated using the stencil-based technique is shown in Figure 8.a.

VI. Hole Filling

To construct the surface of the patient's chest for calculating the corresponding tidal volume, the resulting surface mesh must form a water-tight model. To obtain this model, all occluded and clipped cross-sections must be filled with valid estimates of the surface curvature to form an enclosed volume. These regions are formed by the lack of any surface information about the patient's back and the clipped regions that are not visible to any depth scanning device (*e.g.* cross-sections of the waist, neck, arms). This section describes the process of encapsulating the unbounded region defined by the clipped depth-cloud that defines the patient's chest surface.

A. Planar Hole Fill Algorithm

The clipped regions of the patient's chest provides four primary holes that must be properly filled to enclose the monitored chest volume. Based on the premise of planar grid projection we can easily fill a planar region within an n-sided polygon with a uniform grid of oriented points. This process is used once the edge points of the chest region have been identified and specific joints from the skeleton are used to identify the closest points to the clipped regions from the edge point sets. This is accomplished using the following algorithm: (1) Planar projection of chest edge points $C_p(t)$, (2) 2D Convex Hull on $C_p(t)$, (3) Grid Generation based on AABB of Convex Hull, (4) Point-in-polygon test for included grid points, (5) Generate uniform surface normals.

B. Clip-region Surface Filling

During the process of identifying the chest region of the depth image using the skeletal information, the clipping region of the cylinder introduces newly opened regions that must be filled to construct the chest iso-surface. These regions include the neck, waist, and arms. For the larger clipped neck and waist regions, the characteristic function of the generated surface will be unbounded in these regions and for consistency we cannot allow an arbitrary interpolation scheme to dictate the surface closure in these regions.

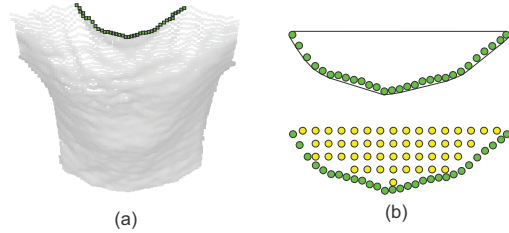(a)   (b)

Fig. 8. Neck-edge points determined by a radial search from the neck joint position (a). The application of the planar hole fill algorithm within the calculated convex hull providing a uniformly closed clip region (b).

Due to this, the planar hole filling algorithm is employed to populate these empty regions with uniformly spaced generated point samples. For each of the generated samples within these regions we assume uniform normals that complement the surface direction required for constructing a iso-volume of the chest region. The image in Figure 8 illustrates this process.

C. Back-region Surface Filling

The remaining hole, caused by the occlusion of the patient, is the completely occluded back region. To ensure consistency of the unknown back surface, we introduce a simple back-fill algorithm to ensure that the naturally occluded region of the back is populated with an estimate of an appropriate surface. This is obtained by utilizing the orientation of the skeletal data (illustrated in Figure 5.b as $\hat{b}$) and projecting all of the existing chest surface points to a backward facing plane with offset from the spine α. The offset value of α only has to provide a consistent distance to the back plane and is defined as $\alpha = 5.0[cm]$ to simply define an average patient back depth. This is an inherently fast approach and provides a closely matched region of points that complement the curvature of the chest points, which aides in the reconstruction of the chest surface.

VII. Surface Reconstruction

The premise of our technique is based on the accurate calculation of a total patients chest volume based on the surface describing the left thorax, right thorax, and abdominal region during the respiration process. To achieve this we utilize an iso-surface reconstruction technique that allows us to efficiently generate a bounded region as volumetric mesh that corresponds to an estimation of the patients tidal volume as the reconstructed model deforms over time. Accurately estimating the tidal volume and respiratory rate using the proposed omni-directional surface technique requires a robust methodology

for surface reconstruction based on a set of unordered, oriented surface points. Additionally the reconstructed surface must maintain the following properties: (1) the generated surface forms a manifold mesh, (2) the triangulation is *water-tight*, and (3) the ordering of every triangle within the surface is consistent. From the premise of extracting a surface from a set of unordered, oriented points, [23] provide an effective means of generating a surface conforming triangulation through the use a variation of the Marching Cubes algorithm [24]. These techniques are consolidated within our model presented in Section III to ensure the construction of a water-tight, manifold mesh with consistent ordering.

In each individual frame recorded from the monitoring device, the surface of the chest is clipped and the corresponding surface normals are estimated and the remaining holes within the surface are closed using our uniform projection technique. Each of these independently acquired oriented point sets are then consolidated into an individual unordered, oriented point cloud. This cloud is then used as the input to the surface generation algorithm. Succinctly, the surface generation process is as follows: (1) the oriented point sets are *splatted* into a voxel grid, (2) the voxel grid is convolved with an integration filter, an estimation of the characteristic function using Fast Fourier coefficients extracted using *FFTW* [25] and (3) the extraction of the surface is achieved using a variant of the marching cubes algorithm with cubic interpolation. The images in Figure 9 illustrate the surface reconstruction process for three individual states during a patient's respiration process.
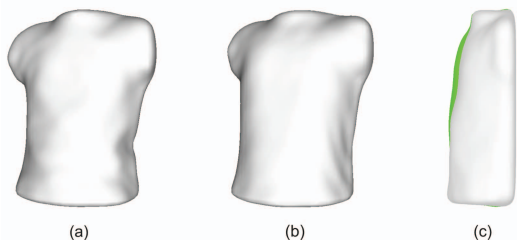

Fig. 9.   Chest surface reconstruction for two independent states: (a) inhale state and (b) inhale state. Even while wearing a normal shirt, the deformation patterns of the patient's chest are visible. As the clothing becomes more form-fitting we naturally see a better representation of the patient's chest rather than the surface of the clothing. The image in (c) illustrates the highlighted cross-sectional difference between the inhale and exhale states.

From the water-tight manifold mesh that is generated through this process, we can simply calculate the volume of this volumetric mesh using the signed tetrahedral volume algorithm. As the resolution of the mesh is decreased, the sample rate increases, however this reduces the accuracy of this technique due to the loss of deformation behavior over the surface of the chest. Similarly, increasing the resolution provides diminishing returns with respect to the accuracy of the estimated tidal volume. Therefore, we select a voxel grid size that provides an accurate chest surface representation.

Note: We do not filter sliver triangles generated from the Marching Cubes surface extraction. This requires an additional process that does not drastically effect our volume calculation.

VIII. Enhanced Tidal Volume Estimation

In this section, we design an algorithm to robustly demodulate fine-grained tidal-volume estimated from volume estimated by the depth-imaging device. Since our method is built on a physiological premise of the harmonic movement between the omni-directional chest expansion and the associated tidal-volume, we utilize this phenomenon as the leading principle for our training algorithm.

A. Training Algorithm

The proposed training process quantifies the relationship between chest movement (mesh volume) and breathing volume of the patient and is only needed once for each patient. During this process, the patient is asked to stand within the device FOV and breathes normally into a spirometer (Figure 1). The ground-truth breathing volume of the patient is recorded by spirometer $V_s(t)$. The main objective is to find a non-linear correlation function F_c of $V_m(t)$ and $V_s(t)$.

Several challenges need to be addressed to properly extract the correlation function F_c. First, the deformations imposed during the patient's breathing cycle are minimal. Second, it is difficult to detect minute volume changes given the noise inherently introducing by body movements and the variation of chest surface caused by the patient's clothing. Finally, the nonuniform movement of the chest during breathing cycle makes the correlation between mesh vs. breathing volume to be dependent on proper experimental setup and the distance of the patient to the monitoring device.
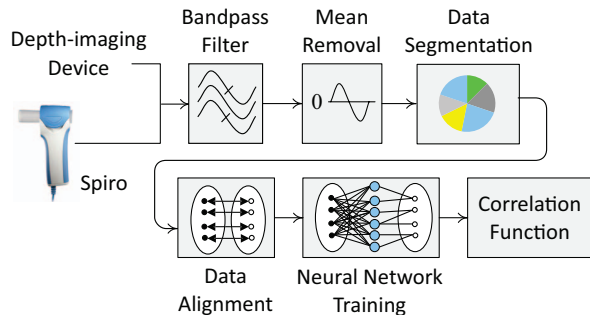

Fig. 10.   The procedure of training process to obtain non-linear correlation function between mesh volume estimated by camera and actual breathing volume collected by ground-truth device (spirometer).

To overcome the these challenges, we exploit the regularity and quasi-periodic nature of a patient's chest movements. In particular, the change in the mesh surface is highly likely to expand and collapse along the same directions illustrated in our omni-directional model. Moreover, the movement direction only changes when the subject changes from inhale to exhale states. Thus, we identify and group chest mesh deformations within one half of a breathing cycle for breathing volume estimation for which per-sample breathing volume is inferred. In the monitoring of this process, we identified that the depth-image noise has a Gaussian distribution. Therefore, noise can be removed by standard filtering mechanisms. Based on the characteristics of the required signal-processing, we

introduce a filtered half-cycle segmentation volume estimation technique based on the 3D measured volume and the ground-truth provided by a spirometer provided as inputs to the proposed training method.

To reduce the noise impact from the 3D measured volume we introduce a bandpass filter based on a frequency of 0.23 $[Hz]$ based on adult [26] respiration rates. Therefore, this define the parameters of this filter with f_{Low} = 0.1$[Hz]$, f_{High} = 1.0$[Hz]$. We also use a low-pass filter with cutoff frequency of 1.0$[Hz]$ to removal non-breathing interference. Due to the effect from the noise especially the distance change between the patient and the monitoring device, the mesh volume changes detected by the device will incur a non-zero mean as the baseline fluctuates unpredictably. To solve this problem, the key idea is to find the noise frequency and remove that frequency component. In our system, we use Savitzky-Golay filter with window size of 5 seconds and order 3 to obtain reliable results that we utilized within our results and evaluation.

Algorithm 2: Training algorithm

---

input : $V_m(t)$
      $V_s(t)$ /* Spirometer */
output: F_c - Correlation function of $V_m(t)$ and $V_s(t)$
if $bMonitoring$ **then**
    Filter $V_m(t)$ (Band-pass filter with cut-off frequencies)
    Mean removal $V_m(t)$

    Align staring point of $V_m(t)$ and $V_s(t)$
    Segment $V_m(t)$ and $V_s(t)$ into n equal segments
    Re-sample $V_s(t)$
    Align $V_m(t)$ and $V_s(t)$)

    Obtain a non-linear correlation function F_c of $V_m(t)$ and
      $V_s(t)$ using neural network.
end
return F_c

The filtered samples are then divided into segments. The segmentation is based on the fact that the breathing activity makes both mesh volume and actual volume data pass the observed baseline repeatedly. After mean removal, the base line is a zero-mean line and the number of inhale and exhale is equal to the number of cross zero line of the captured data. The zero-cross point is then considered as relative referenced points to align both the spirometer and measured volume data to establish the correspondence between the two signals. This provides the basis input for our training procedure. This presented in Algorithm 2. In our method we use a simple $bMonitoring$ that is considered as the start signal when the patient's skeleton is recognized. Once this flag is set, we impose a 5$[s]$ delay for the patient to prepare for the monitoring process.

B. Neural-Network Mesh-to-Volume Correspondence

The Bayesian back-propagation learning algorithm [27] is employed to obtain the correlation of the mesh volume changes over time with the corresponding ground-truth volume. The mesh volume $V_m(t)$ is passed through the system in the first layer of the neural network. Hidden layers are expected to generate non-linear correlation function so that the breathing volume produced from the last layer is as close to the ground truth volume, $V_s(t)$, as possible. To reduce the error between the output volume and the ground truth, the weight of each layer must be determined. We apply the Mackay and Neal [27], [28] weight algorithm for the correlation function. We use sigmoid function, $i.e.$, tanh $S(t) = \frac{1}{1+e^{-t}}$, as the activation function, progressing the number of learning iterations to 1000, or the threshold limit of 0.005 liters.

IX. RESULTS

The results we present are categorized into two sections: (1) technique evaluation and (2) performance of our real-time system. This is due to the implementation of this technique and the potential limitations of the hardware employed in our solution to achieve a real-time estimation. Based on the objective of our approach, we have optimized the performance of our proposed methodology with respect to computation time and tidal volume estimation based on the limitations imposed by the Kinect-2 depth-image acquisition rate with sampling. Furthermore, we illustrate that through the reduction in computational costs within our approach, we are able to extract a highly accurate estimation of the patient's tidal volume at distance range of 1.25$[m]$ to 1.5$[m]$.

A. Tidal-volume Estimation

1) Setup: We conduct an experiment over 4 graduate students (1 female, 3 males) to evaluate the performance of the proposed volume estimation algorithm. The participants are required to stand in front of the camera while using their month to breathe into a spirometer [29]. The participants are not allowed to breath through their nose to ensure the amount of air of inhaled and exhaled are correctly captured by the spirometer. The distance from user is varies from 1.25$[m]$ to 1.75$[m]$. The delay from the time that the skeleton is detected until we collect the data is controllable by the implementation and we found that 5 seconds is a suitable time limit. We also observed that the participant may become tired and breathe abnormally or uncomfortably after 20$[s]$, hence, we break down the data collection process into individual trials (20$[s]$ each). This process is repeated 20 times.

2) Basic volume estimation: The correspondence between the deformations observed in the patient's chest and the estimated tidal volume has been established based on the relationship between the calculated mesh volume and the spirometer ground-truth volume. Inferring the tidal volume of the patient based on our approach allows us to accurately correlate chest deformations with the patient's actual tidal volume. Table I provides an overview of our experimental participants contributing to our evaluated results.

The resulting data-sets are divided into two sets, one use for training, another one is used for evaluation. This presents the results of estimating the tidal volume using our technique for four participants where h is the height, w is weight, cs is chest size, and $error$ is the mean error (based on a 0.2[s]

TABLE I
VOLUME ESTIMATION RESULTS ACROSS PARTICIPANTS

User	sex	age	$h[cm]$	$w[kg]$	$cs[cm]$	$error[l]$
P1	female	28	156	47	35	0.079
P2	male	27	168	70	42	0.075
P3	male	26	170	65	40	0.067
P4	male	24	169	67	41	0.055

window). Using the proposed approach we obtain 92.2% to 94.19% accuracy within our tidal volume estimation with a corresponding 0.055[l] to 0.079[l] error. Figure 11 provides a plot of a representative tidal volume estimation of P2.
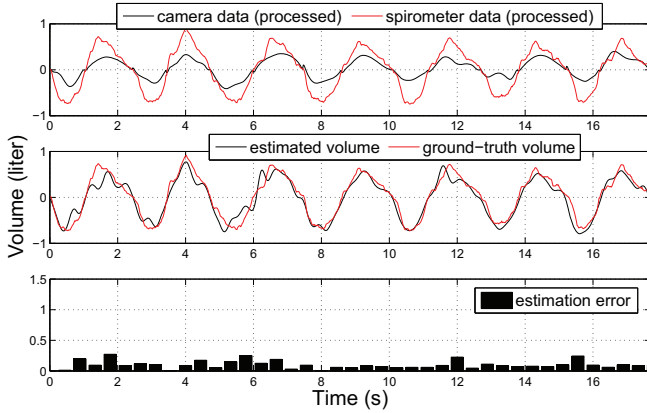

Fig. 11. An example illustrating the processed camera and spirometer correlations. The surface-based estimated tidal volume (top), training-based estimation result (center), and its associated estimation error (bottom).

3) Distance Impact: A critical aspect of using depth-based imaging relates to the effective distance of the monitoring device. The noise incurred due to larger distances will introduce errors and reduces the performance of the surface reconstruction process. We have conducted experiments to evaluate the performance of estimation when varying the distance from camera from 1.25[m] to 1.75[m]. During the process, the student is required to stand in front of the camera and breath through a spirometer when varying the distance between their chest and the camera between each experiment. Figure 12 shows the error distribution over different distances over 10 experiments (20[s] each).
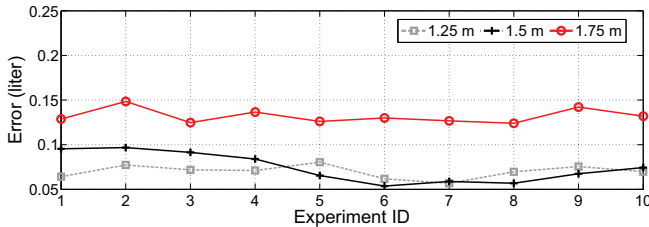

Fig. 12. The depth measurement errors as contributed to the reconstructed surface model. Larger distances provide larger fluctuations in depth measurements, incurring the reduction in accuracy of the estimated tidal volume.

As can be seen from the figure, the system achieves the best performance at the distance of 1.25[m] and the worst performance with the distance of 1.75[m]. As illustrated within the figure, the performance of estimation is reduced

up to 85% (error is approximately 0.15[l]) when the distance increases to 1.75[m].

4) Medical significance: Through the performed experiments, we found that the breathing volume waveforms also represent unique patterns of a participant, which can contribute to clinical analysis of the patient's condition. Figure 13 shows the waveforms of the breathing volumes estimated for four different participants. As can be seen in the figure, the signals (of different participants) are not only different in frequency and amplitude but also represent unique breathing form characteristics. This information is not be obtained by existing state of the art rate estimation techniques.
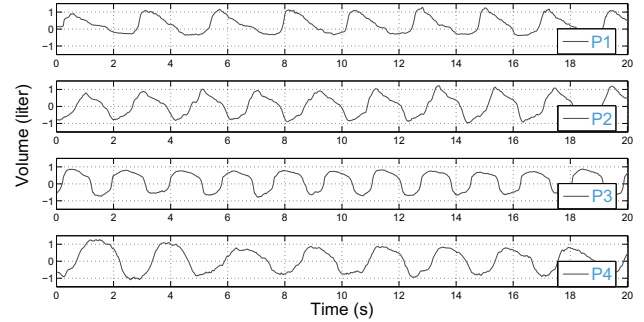

Fig. 13. The tidal volume waveforms of participants $P1 \rightarrow P4$ exhibiting breathing characteristics that uniquely identify their breathing patterns.

Additionally, the results obtained above are independent of the lighting conditions of the patient's environment. Since the device operates off of laser-based depth-images, lighting does not contribute to the requirements for this proposed method.

B. Performance

In the analysis of our proposed tidal volume estimation technique, we incorporate the computation time required to facilitate the construction of the patients chest along with the corresponding volume. The performance results provided in this section illustrate the results of the optimizations implemented within our technique to make the proposed real-time monitoring process possible. This is due to the several steps that are required in our chest surface reconstruction process that are computationally expensive within the real-time domain. Through the optimization of our technique utilizing performance profilers, we are able to obtain a real-time respiratory monitoring system using widely available hardware.

Employing an infrared-based depth-imaging technique, our approach is subject to the distance measurement errors and fluctuations that are naturally imposed by using this form of depth sampling. Additionally, our approach does not use an orthogonal projection of the depth-image to generate the associated depth-cloud, thus the number of samples we collect on the patient's chest varies as a function of distance. The results in Figure 14 illustrate the computation times associated with a patient standing 1.25[m], 1.5[m], and 1.75[m] away from the monitoring device. For each position the number of samples was increased from 1 to 100. When the patient is closer, depth-cloud density rises, giving a more accurate estimation of the chest surface.
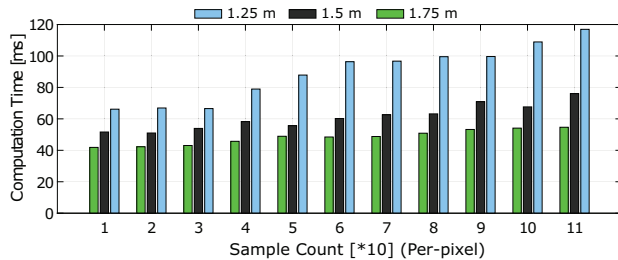
Fig. 14. Computation time of each frame as a function of the number of samples and distance. The experiment was performed at three distances: $1.25[m]$, $1.5[m]$, and $1.75[m]$. For each distance, the number of samples was increased from 1 to 100. At closer distances ($1.25[m]$), higher sampling drastically increases frame computation time.

The performance characteristics of our approach are formed through the four most computationally expensive states. This includes: (1) depth-image sampling with clipping (*Kinect-2 with only depth data*) $47.77[ms]$, (2) chest surface normal estimation $9.51[ms]$, (3) hole filling $1.39[ms]$, and (4) surface reconstruction $19.73[ms]$. Due to the inherent inconsistencies in the depth values provided by the Kinect-2, averaged (smoothed) samples are required to effectively eliminate these natural fluctuations. Based on the minimization of these depth measurement errors obtained by averaging several samples per frame, this sampling obtains the largest portion of the frame computation time. Thus the proposed method is currently only limited by the ability to rapidly sample the patient's chest given the sampling rate of the device.

X. DISCUSSION AND CONCLUSION

In this paper we have presented a methodology of extracting an omni-directional deformation model from the depth and skeletal information provided by depth-imaging devices to provide an accurate estimation of a patients tidal volume given the correlation between the patients chest deformation behavior and their corresponding tidal volume. We have outlined the process of reconstructing an accurate three-dimensional space-time model of a patients chest for the evaluation of the deformations that contribute to an accurate estimate of a patients tidal volume during normal breathing in real-time. The results we present illustrate that through this process we can achieve a highly accurate fine-grain real-time estimation accuracy of 92.2% to 94.19% in the tidal volume estimates of each participant.

The proposed two phase methodology that we have implemented provides an effective means of estimating a patient's tidal volume using a non-contact monitoring solution based on direct tidal volume estimation and improved using patient-specific training. While this technique provides accurate results under the provided set of assumptions, additional circumstances need to be addressed regarding (1) the distance to the patient from the device, (2) depth measurement errors, (3) the patient's posture, (4) the minimization of the patient's movement during the monitoring process and (5) the simplification of the per-patient training data representation. These factors all contribute to the error within the presented methodology and are left to our continued work.

REFERENCES

[1] A. R. Fekr *et al.*, "Design and evaluation of an intelligent remote tidal volume variability monitoring system in e-health applications," in *IEEE J. of Biomedical and Health Informatics*, 2015.
[2] M.-C. Yu *et al.*, "Noncontact respiratory measurement of volume change using depth camera," *IEEE EMBS*, 2012.
[3] Y. Mizobe *et al.*, "Proposal on nonrestraint pulmonary function test using active 3d measurement for body surface," in *World Congress on Medical Physics and Biomedical Engineering*, vol. 2, 2006, pp. 849–852.
[4] C. Li *et al.*, "Accurate doppler radar noncontact vital sign detection using the relax algorithm," *Instrumentation and Measurement, IEEE Trans. on*, vol. 59, no. 3, pp. 687–695, March 2010.
[5] A. Bulanova *et al.*, "The analysis of breath air by laser spectroscopy method for diagnosis of copd," *European Respiratory Journal*, 2014.
[6] L. Boccanfuso and J. M. O'Kane, "Remote measurement of breathing rate in real time using a high precision, single-point infrared temperature sensor," in *IEEE RAS EMBS BioRob*, 2012, pp. 1704–1709.
[7] S. Levine *et al.*, "Use of a triaxial magnetometer for respiratory measurements," *J. Appl. Physio.*, vol. 70, no. 5, pp. 2311–2321, 1991.
[8] P. Dehkordi *et al.*, "Validation of respiratory signal derived from suprasternal notch acceleration for sleep apnea detection," in *IEEE EMBC*, 2011.
[9] T. Reinvuo *et al.*, "Measurement of respiratory rate with high-resolution accelerometer and emfit pressure sensor," in *IEEE Sensors Applications Symposium*, 2006, pp. 192–195.
[10] A. Loblaw *et al.*, "Remote respiratory sensing with an infrared camera using the kinect infrared projector," in *Int. Conference on Image Processing (IPCV)*, 2013.
[11] K. Tan *et al.*, "Real-time vision based respiration monitoring system," in *Int Symposium on Comm. Systems Networks and Digital Signal Processing (CSNDSP)*, 2010, pp. 770–774.
[12] J. Shotton *et al.*, "Efficient human pose estimation from single depth images," in *Decision Forests for Computer Vision and Medical Image Analysis*, 2013, pp. 175–192.
[13] L. Xia *et al.*, "Human detection using depth information by kinect," in *IEEE Computer Vision and Pattern Recognition Workshops*, June 2011, pp. 15–22.
[14] S. Gumhold *et al.*, "Feature extraction from point clouds," in *In Proc. of the Int. Meshing Roundtable*, 2001, pp. 293–305.
[15] M. Al-Mardini *et al.*, "Classifying obstructive sleep apnea using smartphones," vol. 52, pp. 251–259, 2014.
[16] P. Nguyen *et al.*, "Poster: Continuous and fine-grained respiration volume monitoring using continuous wave radar," in *ACM Mobicom*, 2015, pp. 266–268.
[17] H. Aoki *et al.*, "Non-contact respiration measurement using structured light 3-d sensor," in *SICE Annual Conference*, Aug 2012, pp. 614 – 618.
[18] M. Kazhdan, "Reconstruction of solid models from oriented point sets," in *Proceedings of the Third Eurographics Symposium on Geometry Processing*. Eurographics Association, 2005.
[19] C. Zhang and T. Chen, "Efficient feature extraction for 2d/3d objects in mesh representation," in *Image Processing, Int. Conference on*, vol. 3, 2001, pp. 935–938 vol.3.
[20] N. J. Mitra *et al.*, "Estimating surface normals in noisy point cloud data," *International Journal of Computational Geometry and Applications*, 2004.
[21] R. B. Rusu *et al.*, "Towards 3d point cloud based object maps for household environments," *Robotics and Autonomous Systems Journal (Special Issue on Semantic Knowledge)*, 2008 2008.
[22] S. Holzer *et al.*, "Adaptive neighborhood selection for real-time surface normal estimation from organized point cloud data using integral images," *Int. Conference on Intelligent Robots and Systems*, 2012.
[23] H. Hoppe *et al.*, "Surface reconstruction from unorganized points," in *SIGGRAPH*. ACM, 1992, pp. 71–78.
[24] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," in *SIGGRAPH*. ACM, 1987.
[25] M. Frigo and S. G. Johnson. (1998) Fastest fourier transform in the west. [Online]. Available: http://www.fftw.org/
[26] P. Sebel, "Respiration, the breath of life," 1985.
[27] D. MacKay, "A practical bayesian framework for backpropagation networks," vol. 4, no. 3, pp. 448–472, 1992.
[28] R. M. Neal, "Bayesian learning for neural networks," 1996.
[29] Vernier. Spirometer SBR-BTA. [Online]. Available: http://www.vernier.com/products/sensors/spr-bta/