©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-045

Im2Fit: Fast 3D Model Fitting and Anthropometrics using Single Consumer Depth Camera and Synthetic Data

Qiaosong Wang [1]***, Vignesh Jagadeesh*** [2]***, Bryan Ressler*** [2] ***and Robinson Piramuthu*** [2]
[1]University of Delaware, 18 Amstel Ave, Newark, DE 19716
[2]eBay Research Labs, 2065 Hamilton Avenue, San Jose, CA 95125

Abstract

Recent advances in consumer depth sensors have created many opportunities for human body measurement and modeling. Estimation of 3D body shape is particularly useful for fashion e-commerce applications such as virtual try-on or fit personalization. In this paper, we propose a method for capturing accurate human body shape and anthropometrics from a single consumer grade depth sensor. We first generate a large dataset of synthetic 3D human body models using real-world body size distributions. Next, we estimate key body measurements from a single monocular depth image. We combine body measurement estimates with local geometry features around key joint positions to form a robust multi-dimensional feature vector. This allows us to conduct a fast nearest-neighbor search to every sample in the dataset and return the closest one. Compared to existing methods, our approach is able to predict accurate full body parameters from a partial view using measurement parameters learned from the synthetic dataset. Furthermore, our system is capable of generating 3D human mesh models in real-time, which is significantly faster than methods which attempt to model shape and pose deformations. To validate the efficiency and applicability of our system, we collected a dataset that contains frontal and back scans of 83 clothed people with ground truth height and weight. Experiments on real-world dataset show that the proposed method can achieve real-time performance with competing results achieving an average error of 1.9 cm in estimated measurements.

Introduction

Recent advances in 3D modeling and depth estimation have created many opportunities for non-invasive human body measurement and modeling. Particularly, getting precise size and fit recommendation data from consumer depth cameras such as Microsoft® Kinect™ device can reduce returns and improve user experience for online fashion shopping. Also, applications like virtual try-on or 3D personal avatars can help shoppers visualize how clothes would look like on them. According to [1, 2], about $44.7B worth of clothing was purchased online in 2013. Yet, there was an average return rate of 25% because people could not physically try the clothes on. Therefore, an accurate 3D model of the human body is needed to provide accurate sizing and measurement information to guide online fashion shopping. One standard approach incorporates light detection and ranging (LiDAR) data for body scanning [3, 4]. An alternative approach is to use a calibrated multi-camera rig and reconstruct the 3D human body model using structure from motion (SfM) and Multi-View Stereo (MVS) [5], which generates impressive results when combined with motion capture technology [6]. A more recent work [7] aims at estimating high quality 3D models by using high resolution 3D human scans during registration process to get statistical model of shape deformations. Such scans in the training set are very limited in variations, expensive and cumbersome to collect and nevertheless do not capture the diversity of the body shape over the entire population. All of these systems are bulky, expensive, and require expertise to operate.

Recently, consumer grade depth camera has proven practical and quickly progressed into markets. These sensors cost around $200 and can be used conveniently in a living room. Depth sensors can also be integrated into mobile devices such as tablets [8], cellphones [9] and wearable devices [10]. Thus, depth data can be obtained from average users and accurate measurements can be estimated. However, it is challenging to produce high-quality 3D human body models, since such sensors only provide a low-resolution depth map (typically 320 × 240) with a high noise level.

In this paper, we propose a method to predict accurate body parameters and generate 3D human mesh models from a single depth map. We first create a large synthetic 3D human body model dataset using real-world body size distributions. Next, we extract body measurements from a single frontal-view depth map using joint location information provided by OpenNI [11]. We combine estimates of body measurements with local geometry features around joint locations to form a robust multi-dimensional feature vector. This allows us to conduct a fast nearest-neighbor search to every 3D model in the synthetic dataset and return the closest one. Since the retrieved 3D model is fully parameterized and rigged, we can easily generate data such as standard full body measurements,labeled body parts, etc. Furthermore,we can animate the model by mapping the 3D model skeleton to joints provided by Kinect™. Given the shape and pose parameters describing the body, we also developed an approach to fit a garment to the model with realistic wrinkles. A key benefit of our approach is that we only calculate simple features from the input data and search for the closest match in the highly-realistic synthetic dataset. This allows us to estimate a 3D body avatar in real-time, which is crucial for practical virtual reality applications. Also, compared to real-world human body datasets such as CAESAR [12] and SCAPE [13], the flexibility of creating synthetic models enables us to represent more variations on body parameter distributions, while lowering cost and refraining from legal privacy issues involving human subjects. In summary, we make the following contributions:

- We build a complete software system for body measurement extraction and 3D human model creation. The system

DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-045

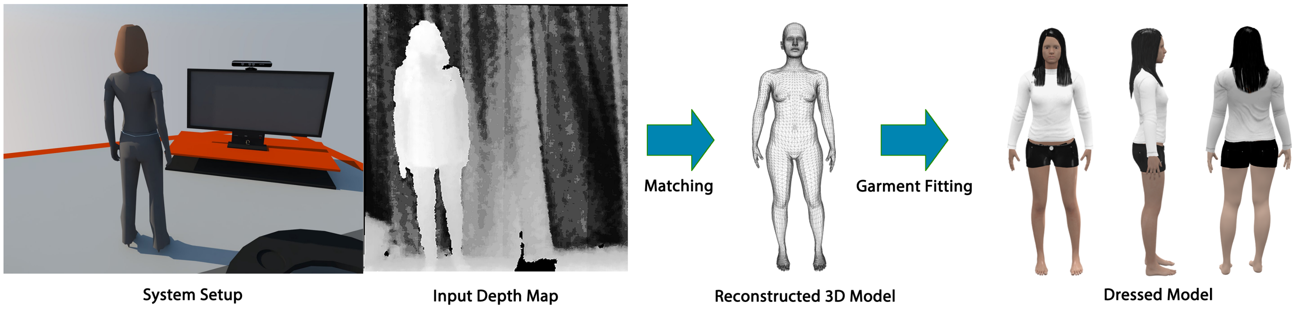


Figure 1: **Im2Fit**. We propose a system that uses a single consumer depth camera to (i) estimate 3D shape of human, (ii) estimate key measurements for clothing and (iii) suggest clothing size. Unlike competing approaches that require a complex rig, our system is very simple and fast. It can be used at the convenience of existing set up of living room.

only requires a single input depth map with real-time performance.

- We propose a method to generate a large synthetic human dataset following real-world body parameter distributions. This dataset contains detailed information about each sample, such as body parameters, labeled body parts and OpenNI-compatible skeletons. To our knowledge, we are the first to use large synthetic data that match distribution of true population for the purpose of model fitting and anthropometrics.
- We design a robust multi-dimensional feature vector and corresponding matching schemes for fast 3D model retrieval. Experiments show that our proposed method that uses simple computations on the input depth map returns satisfactory results.

The remainder of this paper is organized as follows. Section gives a summary of previous related work. Section discusses Im2Fit - our proposed system. Section shows our experimental results. Section provides some final conclusions and directions for future work.

Background and Related Work

We follow the pipeline of generating a large synthetic dataset, and then perform feature matching to obtain the corresponding 3D human model. A wealth of previous work has studied these two problems, and we mention some of them here to contrast with our approach.

3D Human Body Datasets: Despite vast literature concerning depth map based human body modeling, only a limited number of datasets are available for testing. These datasets typically contain real-world data to model the variation of human shape, and require a license to purchase.

The CAESAR [12] dataset contains few thousand laser scans of bodies of volunteers aged from 18 to 65 in the United States and Europe. The raw data of each sample consists of four scans from different viewpoints. The raw scans are stitched together into a single mesh with texture information. However, due to occlusion, noise and registration errors, the final mesh model is not complete.

The SCAPE [13] dataset has been widely used in human shape and pose estimation studies. Researchers found it useful to reshape the human body to fit 3D point cloud data or 2D human silhouettes in images and videos. It learns a pose deformation model from 71 registered meshes of a single subject in various poses. However, the human shape model is learned from different subjects with a standard pose. The main limitation of the SCAPE model is that the pose deformation model is shared by different individuals. The final deformed mesh is not accurate, since the pose deformation model is person-dependent. Also, the meshes are registered using only geometric information, which is unreliable due to shape ambiguities.

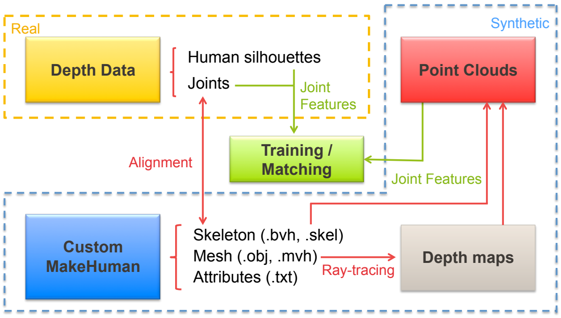


Figure 2: **Block Diagram**. This illustrates the interactions between various components of Im2Fit, along with relevant data. Details are in Section .

Recently, Bogo *et al.* [14] introduced a new dataset called FAUST which contains 300 scans of 10 people with different poses. The meshes are captured by multiple stereo cameras and speckle projectors. All subjects are painted with high frequency textures so that the alignment quality can be verified by both geometry and appearance information.

Human Shape Estimation: A number of methods have been proposed to estimate human shape based on raw 3D data generated by a depth sensor or multiple images. Early work by Blanz and Vetter [15] use dense surface and color data to model face shapes. They introduced the term *3D morphable model* to describe the idea of using a single generic model to fit all real-world face shapes. In their work, pose and shape are learned separately and then combined by using linear blend skinning (LBS). This inspires the SCAPE [13] model, which models deformation as combination of a pose transformation learned from a single person with multiple poses and a shape transformation learned from multiple people with a standard pose. Balan *et al.* [16] fit the SCAPE model to multiple images. They later extend their work to estimate body shape under clothing [17]. Guan *et al.* [18] estimated

Table 1: **Demographics.** Distribution of height (cm) and weight (kg) for selected age and sex groups (mean ± 1 std dev). The age and sex composition are obtained by using the 2010 census data [23], while the height and weight distributions are obtained by using the NHANES 1999-2002 census data [24]. This table was used to generate synthetic models using MakeHuman [25]. Note that, unlike our approach, datasets such as CAESAR [12] do not truly encompass the diversity of the population and are cumbersome and expensive to collect.

Age	Gender			
	Male		Female	
	Height	Weight	Height	Weight
18-24	176.7±0.3	83.4±0.7	162.8±0.3	71.1±0.9
25-44	176.8±0.3	87.6±0.8	163.2±0.3	75.3±1.0
45-64	175.8±0.3	88.8±0.9	162.3±0.3	76.9±1.1
65-74	174.4±0.3	87.1±0.6	160.0±0.2	74.9±0.6

human shape and pose from a single image by optimizing the SCAPE model using a variety of cues such as silhouette overlap, edge distance and smooth shading. Hasler *et al.* [19] developed a method to estimate 3D body shapes under clothing by fitting the SCAPE model to images using ICP [20] registration and Laplacian mesh deformation. Weiss [21] proposed a method to model 3D human bodies from noisy range images from a commodity sensor by combining the silhouette overlap term, prediction error and an optional temporal pose similarity prior. More recently, researchers tackle the problem of decoupling shape and pose deformations by introducing a tensor based body model which jointly optimizes shape and pose deformations [22].

Im2Fit: The Proposed System

The overall processing pipeline of our software system can be found in Figure 2. Algorithm 1 summarizes the various steps involved to estimate relevant anthropometrics for clothing fitment, once depth data and joint locations have been acquired.

Depth Sensor Data Acquisition

The goal of real data acquisition process is to obtain user point cloud and joint locations, and extract useful features characterizing the 3D shape, such as body measurements and 3D surface feature descriptors. We first extract human silhouette from a Kinect$^{\text{TM}}$ RGB-D frame and turn it into a 3D pointcloud. Next, we obtain joint locations by using OpenNI [11] which provides a binary segmentation mask of the person, along with skeletal keypoints corresponding to different body joints. There are 15 joint positions in 3D real world coordinates (Figure 3(a)). They can be converted to 2D coordinates on the imaging plane by projection.

Depth Sensor Data Processing - Measurements and Feature Extraction

Once we obtain depth map and joint locations, we generate features such as body measurements and Fast Point Feature Histograms (FPFH)[1] [26] around these joints. The body measurements include height, length of arm & leg, girth of neck, chest, waist & hip. These measurements can help us predict important user attributes such as gender, height, weight, clothing size. The aforementioned features are concatenated into a single feature vector for matching the features in the synthetic dataset.

We observe that the 6 points on the torso *(NE, TO, LS, RS, LH, RH)* do not exhibit any relative motion and can be regarded on a whole rigid body. Therefore, we define the principal coordinate axes $\{\mathbf{u},\mathbf{v},\mathbf{w}\}$ as $\mathbf{u} = \frac{\overrightarrow{(NE,TO)}}{\|\overrightarrow{(NE,TO)}\|}, \mathbf{v} = \frac{\overrightarrow{(LS,RS)}}{\|\overrightarrow{(LS,RS)}\|}, \mathbf{w} = \mathbf{u}\times\mathbf{v}$, as long as the following condition is satisfied:

$$\left(\mathbf{u}\cdot\overrightarrow{(TO,\frac{LH+RH}{2})}\Big/\|\overrightarrow{(TO,\frac{LH+RH}{2})}\|\right) < \varepsilon_{\mathbf{1}} \tag{1}$$

where $\overrightarrow{(A,B)}$ is the vector from point A to point B, $\times$ is the vector cross-product, $\cdot$ is the vector dot-product, $\|\cdot\|$ is the Euclidean distance, and ε_1 is a small positive fraction. We used $\varepsilon_1 = 0.1$. Condition (1) ensures that the posture is almost vertical. For simplicity, if this is not satisfied, we treat the data as unusable since error in measurement estimates will be large.

Height: To calculate height, we first extract contour of the segmented 2D human silhouette, which was obtained by thresholding the depth map and projection on to 2D plane defined by **u** and **v**. Next, we pick those points on the contour that satisfy the following condition:

$$\left(\left|\mathbf{v}\cdot\overrightarrow{(TO,P_c)}\right|\Big/\left\|\overrightarrow{(TO,P_c)}\right\|\right) < \varepsilon_2 \tag{2}$$

where P_c is an arbitrary point on the contour and ε_2 is a small positive fraction. We used $\varepsilon_2 = 0.1$. These 2D points lie approximately on **u**. We sort them by their y coordinates [2] and find the top and bottom points. These points are converted to 3D real-world coordinates and the estimated height can be calculated as the Euclidean distance between the two points.

Length of Sleeve and Legs: The sleeve length can be calculated as the average of $\left\|\overrightarrow{(LA,LE)}\right\| + \left\|\overrightarrow{(LE,LS)}\right\| + \left\|\overrightarrow{(LS,NE)}\right\|$ and $\left\|\overrightarrow{(RA,RE)}\right\| + \left\|\overrightarrow{(RE,RS)}\right\| + \left\|\overrightarrow{(RS,NE)}\right\|$. Note that these points are in 3D real-world coordinates. Similarly, length of legs can be obtained as the average of $\left\|\overrightarrow{(LH,LK)}\right\| + \left\|\overrightarrow{(LK,LF)}\right\|$ and $\left\|\overrightarrow{(RH,RK)}\right\| + \left\|\overrightarrow{(RK,RF)}\right\|$.

Girth of neck, shoulder, chest, waist and hip: To get estimates of neck, shoulder, chest, waist and hip girth, we need to first define a 3D point **x**, and then compute the intersection between the 3D point cloud and the plane passing through **x** and perpendicular to **u**. Since the joints tracked by OpenNI are designed to be useful for interactive games, rather than being anatomically precise, we need to make some adjustments to the raw joint locations. Here we define new joint locations as follows:

$$\begin{aligned} \mathbf{x}_{neck} &= \frac{NE+HE}{2}, \mathbf{x}_{shoulder} = \frac{\left(\mathbf{x}_{neck} + \frac{LS+RS}{2}\right)}{2} \\ \mathbf{x}_{chest} &= \frac{NE+TO}{2}, \mathbf{x}_{waist} = TO, \mathbf{x}_{hip} = \frac{LH+RH}{2} \end{aligned} \tag{3}$$

We can only get frontal view of the user, since the user is always directly facing the depth camera and we use a single depth

[1] FPFH features are simple 3D feature descriptors to represent geometry around a specific point and are calculated for all joints.

[2] Note that OpenNI uses GDI convention for the 2D coordinate system.

channel as input. Therefore, we fit an ellipse [27] to the points on every cross-section plane to obtain full-body measurements. Such points are defined as $\left\{\mathbf{p} \in \text{Point Cloud}, s.t. \left|\frac{(\mathbf{x}-\mathbf{p})\cdot\mathbf{u}}{\|\mathbf{x}-\mathbf{p}\|}\right| < \varepsilon_3\right\}$, where ε_3 is a small positive fraction. We used $\varepsilon_3 = 0.1$. Since our method is simple and computationally efficient, the measurements can be estimated in real time.

Algorithm 1: Frontal Measurement Generation

```
Data: Depth Map $\mathbf{D}$ and estimated set of joints from pose
      estimation $\{j_t\}_{t=1,\ldots,T}$
Result: Set of Measurements $\mathbf{M}$, where $t^{th}$ row is the
        measurement around joint $j_t$
while onEndCalibration=failure do
    Call StartPoseDetection ;
    if User detected then
        Draw skeletons;
        Compute principal axes $\{\mathbf{u},\mathbf{v},\mathbf{w}\}$ ;
        Request calibration pose from user;
        Call RequestCalibrationSkeleton;
        if User calibrated then
            Save calibrated pose;
            Call onEndCalibration;
            while onEndCalibration=success do
                Call StartTrackingSkeleton ;
                for Joint label $t = 1 : T$ do
                    if $j_t$ is available then
                        Compute cross section plane.
                         /* defined by v and w passing
                         through joint position $j_t$ */ ;
                        Compute intersection points;
                        Ellipse fitting to obtain $m_t$;
                        if $m_t$ is available for all T
                         joints then
                            Save M;
                            Break;
```

Synthetic Data Generation - Depth Map Rendering and Joint Remapping

For realistic synthetic data generation, the goal is to create a large number of synthetic human data according to real world population distributions and save the feature vectors for each model. In order to generate the synthetic human model dataset, we developed a plug-in[3] for MakeHuman [25].

MakeHuman is an open-source python framework designed to prototype realistic 3D human models. It contains a standard rigged human mesh, and can generate realistic human characters based on normalized attributes for a specific virtual character, namely Age, Gender, Height, Weight, Muscle and Ethnic origin (Caucasian, Asian, African). Our plug-in only modifies 4 attributes: Age, Gender, Height and Weight, which follows the distribution shown in Table 1. This is because we only care about the 3D geometry of the human body. Every synthetic model contains a Wavefront .obj file (3D mesh), a .skel skeleton file (Joint location), a Biovision Hierarchy .bvh (Rigged skeleton data) file and a text file containing the 9 above-mentioned attributes. We tested our plug-in on a desktop computer with Intel Xeon E5507 (4 MB Cache, 2.26 GHz) CPU and 12 GB RAM. The modified software is able to generate around 50,000 synthetic models in 12 hours (roughly a model every second).

Since the synthetic data is clean, parametric and has fully labeled body parts and skeletons, we can match real world feature vectors extracted from depth frames directly to the MakeHuman synthetic feature vectors, and use the MakeHuman 3D parametric mesh model to replace incomplete and noisy point cloud data from consumer grade depth sensor. Since it is straightforward to establish joint correspondences, the 3D synthetic model can be deformed into any pose according to the pose changes of the user. The 3D mesh model can now be used for various applications such as virtual try-on, virtual 3D avatar or animation. Since it is difficult to directly process the MakeHuman mesh data, we also developed a scheme to render the mesh model into frontal and back depth maps and equivalent point clouds. The local 3D features can be extracted easily from these point clouds by using the Point Cloud Library (PCL) [28].

The above process has several advantages. Firstly, our customized plug-in is able to generate a large number of realistic synthetic models with real world age, height, weight, or gender distribution, even dressed models with various clothing styles. Secondly, the feature vectors can be matched in real time. Thus, we can find the matching 3D model as soon as the user is tracked by OpenNI. Compared to methods such as [21, 22, 29] which try to generate parametric human model directly on the raw point cloud, our method is more robust and computationally efficient.

Synthetic Data Retrieval - Online 3D Model Retrieval

Now that we obtained real-world and synthetic data in compatible formats, we can match feature vectors extracted from Kinect$^{\text{TM}}$ depth frames directly to those generated from MakeHuman [25]. The retrieved 3D human model can be used to replace the noisy, unorganized raw Kinect$^{\text{TM}}$ point cloud.

To reach the goal of real-time 3D model retrieval, we need to define a simple yet effective feature vector first. We divide the feature vector into three groups with different weights: global body shape, gender and local body shape.

The first group contains 4 parameters: height and length of sleeve, leg and shoulder. This captures the longer dimensions of the human body.

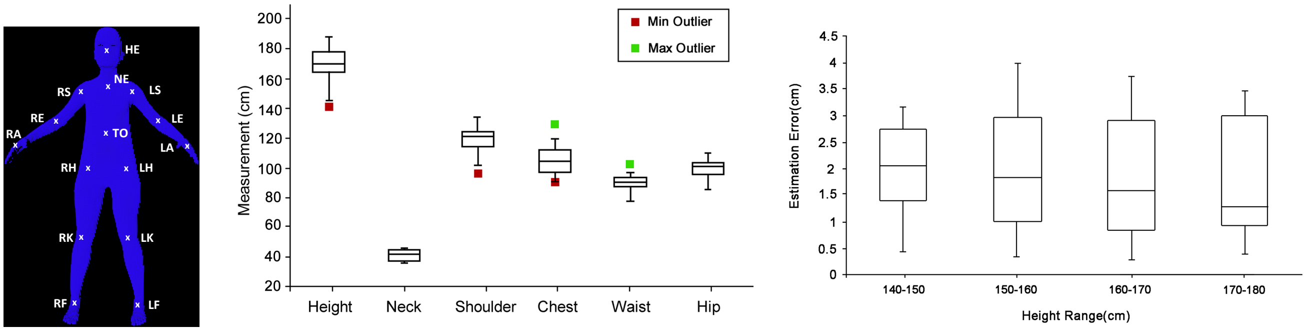
The second group contains two ratios. They are defined as follows:

$$ratio_1 = \frac{\|\widehat{Chest}\|}{\|\overrightarrow{Chest}\|}, ratio_2 = \frac{Hip}{Waist} \qquad (4)$$

where $\|\widehat{Chest}\|$ represents the 3D surface geodesic distance and $\|\overrightarrow{Chest}\|$ represents the Euclidean distance between point LS and RS. If two body shapes are similar, these two ratios tend to be larger for females and smaller for males.

[3] We plan to make this plug-in available for research purpose and hope will flourish research in this area, which is stunted by the complexity of data acquisition for 3D human models and lack of royalty-free large public datasets.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-045



(a) Interface (b) Measurement Distribution (c) Height Estimation Error

Figure 3: **Estimated Anthropometrics**. (a) We use 15 joints head, neck, torso and left/right hand, elbow, shoulder, hip, knee, foot, which we term *HE, NE, TO, LA, RA, LE, RE, LS, RS, LH, RH, LK, RK, LF, RF*, respectively. (b) Distribution of key estimated measurements. We collected depth maps from 83 people and generated height, neck, shoulder, chest, waist and hip measurements for each person. Plotted is the sample minimum, median and maximum, lower and upper quartile and outliers for different measurements. (c) For a given height range, the distribution of height estimation errors are plotted. The overall mean estimation error for 83 samples is 1.87 cm.

The third group contains FPFH (33 dimensions) features [26] computed at all 15 joints with a search radius of 20 centimeters. This describes the local body shape. The feature vector has $4+2+33\times 15 = 501$ dimensions. Any pair of feature vectors can be compared simply by using their L_2 distance.

We use nearest neighbor search to find the closet match in the synthetic dataset. The size of the feature vector dataset for 50,000 synthetic models is roughly 25 MB, and a single query takes about 800 ms to complete on a 2.26 GHz Intel Xeon E5507 machine.

Experiments

We conducted our experiments on clothed subjects with various body shapes. Figure 3(b) shows the distribution of measurements from 83 people predicted by our system. For privacy issues, we were only able to get ground truth height, gender and clothing size data.

The gender predicted by $Ratio_1$ and $Ratio_2$ is correct for all 83 samples. The height estimation error for different height ranges is shown in Figure 3(c). The overall mean estimation error for the 83 samples is 1.87 centimeters. Compared to [21], our system has obtained competitive accuracy with a much simpler algorithm. The results of our system can be qualitatively evaluated from Figure 5. To verify the prediction accuracy, we applied the Iterative Closest Point (ICP) [20] registration algorithm to obtain the mean geometric error between the retrieved 3D mesh model and the raw point cloud data for the four subjects shown in Figure 5. The results are shown in Figure 6. We simply aligned the skeleton joints for ICP initialization. After 6 ICP iterations, the mean registration error converges to below 9 centimeters, for all four subjects. This is because the feature vector gives more weights to frontal measurements. The mesh alignment can be fine tuned by ICP after several iterations. However, the initial joint locations are only accurate on the 2D plane, and may shift in the z-direction due to random noise or holes on the point cloud surface.

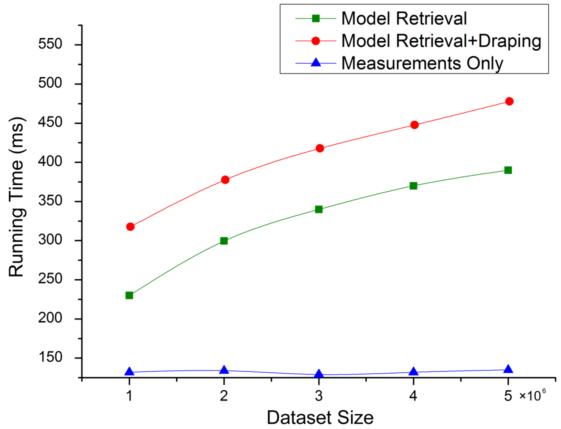


Figure 4: **Runtime**. Runtime of our complete system with increasing size of the synthetic dataset. For synthetic dataset with given number of samples, we evaluate the average runtime end-end, to estimate relevant measurements, naked 3D models and clothed 3D models.

Runtime of 3D model retrieval algorithm with different size of the synthetic dataset is summarized in Figure 4. The algorithm was run on a single desktop without GPU acceleration to obtain relevant measurements, naked body model and clothed model. The measurement estimation only depends on the input depth map. So the running time is almost constant on a depth map with given resolution. Generating clothed models is slower than naked models because of the garment fitting and rendering overhead which can be improved by GPU acceleration. On a dataset containing 5×10^6 samples, we achieved runtime of less than 0.5 seconds for a single query, with little memory usage. Compared to [21] which takes approximately 65 minutes to optimize, our method is significantly faster, while still maintaining competitive accuracy.

Using measurements from the body meshes, we also predicted clothing sizes and compared with ground truth data provided by the participants. The overall clothing size prediction accuracy is 87.5%. We estimated size of T-shirt (*e.g.* XS, S, M, L, XL, 2XL, 3XL).

Conclusion

We presented a system for human body measurement and modeling using a single consumer depth sensor, using only the depth channel. Our method has several advantages over existing methods. First, we propose a method for generating synthetic human body models following real-world body parameter distribu-

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-045

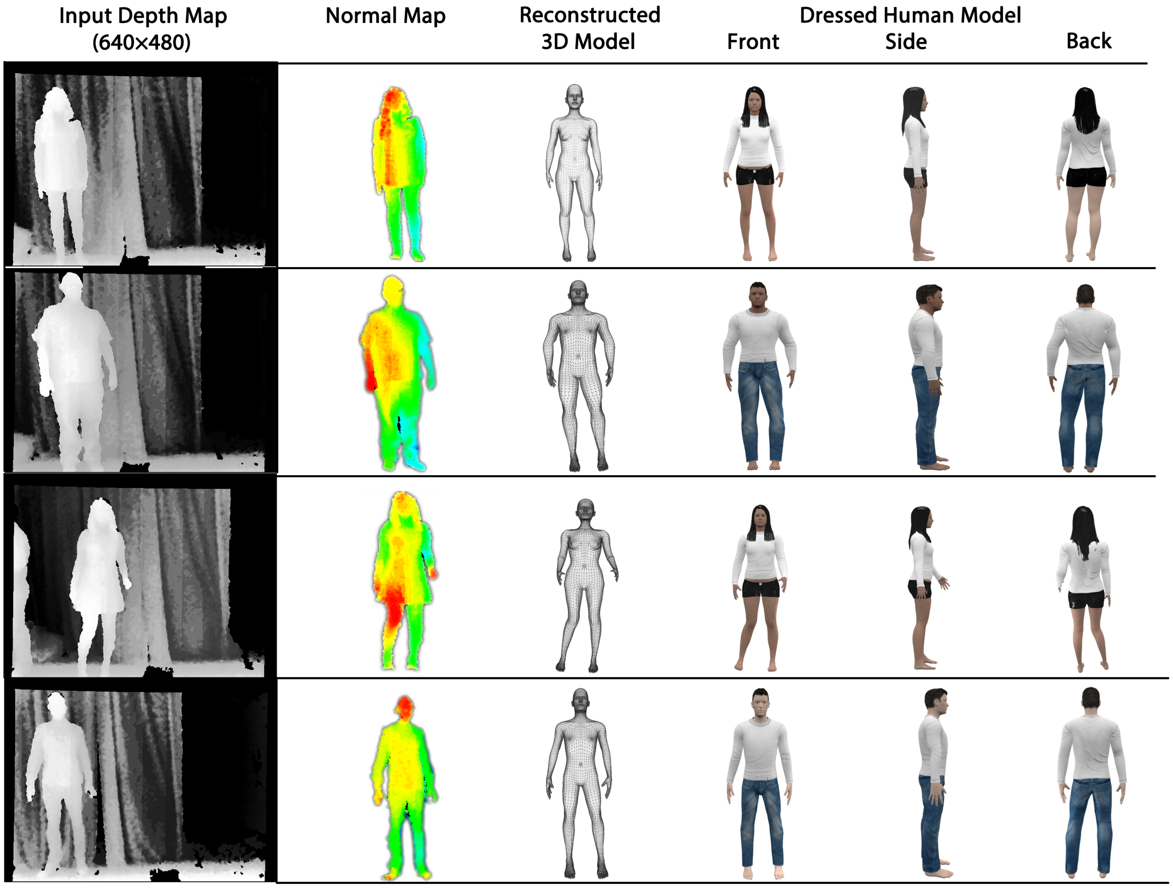


Figure 5: **Qualitative Evaluation**. The first column shows the raw depth map from Kinect^{TM} 360 sensor, the second column shows the normal map which we use to compute the FPFH [26] features. The third column shows the reconstructed 3D mesh model, and the last three columns show the front, side and back view of the 3D model after basic garment fitting.

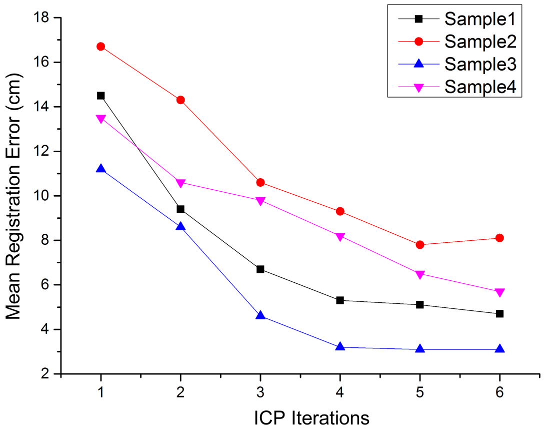


Figure 6: **ICP Error**. Mean registration error per ICP [20] iteration for different samples. We evaluated ICP registration error for 4 samples in the dataset. The raw depth maps and estimated body shapes for the 4 samples can be found in Figure 5.

tions. This allows us to avoid complex data acquisition rig, legal and privacy issues involving human subject research, and at the same time create a synthetic dataset which represents more variations on body shape distributions. Also, models inside the synthetic dataset are complete and clean (*i.e.* no incomplete surface), which is perfect for subsequent applications such as garment fitting or animation. Second, we presented a scheme for real-time 3D model retrieval. Body measurements are extracted and combined with local geometry features around key joint locations to form a robust multi-dimensional feature vector. 3D human models are retrieved by using a fast nearest neighbor search, and the whole process can be done within half a second on a dataset containing 5×10^6 samples. Experiment results have shown that our system is able to generate accurate results in real-time, thus is particularly useful for home-oriented body scanning applications on low computing power devices such as depth-camera enabled smartphones or tablets.

References

[1] eMarketer, “United states apparel and accessories retail e-commerce revenue.” http://www.statista.com/statistics/278890/ (2012 (Accessed on Sepember, 2014)).

[2] Fits.me, “Garment returns and apparel ecommerce: avoidable causes, the impact on business and the simple solution.” fits.me/resources/whitepapers/ (2012 (Accessed on Sepember, 2014)).

[3] “2h3d scanning.” http://www.2h3dscanning.com/ (2014 (Accessed on Sepember, 2014)).

[4] “Gentle giant studios.” http://www.gentlegiantstudios.com/ (2014 (Accessed on

DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-045

Sepember, 2014)).

[5] "Ten 24 digital capture." http://ten24.info/ (2014 (Accessed on Sepember, 2014)).

[6] "Quantic dream studio." http://www.quanticdream.com/ (2014 (Accessed on Sepember, 2014)).

[7] Tsoli, A., Loper, M., and Black, M. J., "Model-based anthropometry: Predicting measurements from 3d human scans in multiple poses," *IEEE Winter Conference on Applications of Computer Vision (WACV), 2014* , 83–90, IEEE, Steamboat Springs, CO, USA (Mar. 2014).

[8] "The structure sensor." http://structure.io (2014 (Accessed on Sepember, 2014)).

[9] "Google project tango." http://www.google.com/atap/projecttango/ (2014 (Accessed on Sepember, 2014)).

[10] "Microsoft hololens." http://www.microsoft.com/microsoft-hololens (2015 (Accessed on Sepember, 2015)).

[11] "Openni sdk." [10]http://structure.io/openni (2014 (Accessed on Sepember, 2014)).

[12] SAE, "Civilian american and european surface anthropometry resource project," (2012).

[13] Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., and Davis, J., "Scape: Shape completion and animation of people," *ACM Trans. Graph.* **24**, 408–416 (July 2005).

[14] Bogo, F., Romero, J., Loper, M., and Black, M. J., "FAUST: Dataset and evaluation for 3D mesh registration," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.* , IEEE, Piscataway, NJ, USA (June 2014).

[15] Blanz, V. and Vetter, T., "A morphable model for the synthesis of 3d faces," *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* , 187–194, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (1999).

[16] Balan, A. O., Sigal, L., Black, M. J., Davis, J. E., and Haussecker, H. W., "Detailed human shape and pose from images," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.* , 1–8, IEEE (2007).

[17] Bălan, A. O. and Black, M. J., "The naked truth: Estimating body shape under clothing," *European Conference on Computer Vision (ECCV),2008* , 15–29, Springer-Verlag, Berlin, Heidelberg (2008).

[18] Guan, P., Weiss, A., Balan, A. O., and Black, M. J., "Estimating human shape and pose from a single image," *IEEE International Conference on Computer Vision (ICCV), 2009* , 1381–1388, IEEE (2009).

[19] Hasler, N., Stoll, C., Rosenhahn, B., Thormahlen, T., and Seidel, H.-P., "Technical section: Estimating body shape of dressed humans," *Comput. Graph.* **33**, 211–216 (June 2009).

[20] Besl, P. J. and McKay, N. D., "Method for registration of 3-d shapes," *Robotics-DL tentative* , 586–606, International Society for Optics and Photonics (1992).

[21] Weiss, A., Hirshberg, D., and Black, M., "Home 3d body scans from noisy image and range data," *IEEE International Conference on Computer Vision (ICCV), 2011* , 1951–1958 (Nov 2011).

[22] Chen, Y., Liu, Z., and Zhang, Z., "Tensor-based human body modeling," *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition* , 105–112, IEEE Computer Society, Washington, DC, USA (2013).

[23] "Age and sex composition 2010 census briefs." http://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf (2014 (Accessed on Sepember, 2014)).

[24] "Mean body weight, height, and body mass index, united states 19602002." http://www.cdc.gov/nchs/data/ad/ad347.pdf (2014 (Accessed on Sepember, 2014)).

[25] "Makehuman: Open source tool for making 3d characters." http://www.makehuman.org/ (2014 (Accessed on Sepember, 2014)).

[26] Rusu, R., Blodow, N., and Beetz, M., "Fast point feature histograms (fpfh) for 3d registration," *IEEE International Conference on Robotics and Automation (ICRA), 2009* , 3212–3217 (May 2009).

[27] Fitzgibbon, A., Pilu, M., and Fisher, R. B., "Direct least square fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(5), 476–480 (1999).

[28] Rusu, R. and Cousins, S., "3d is here: Point cloud library (pcl)," *IEEE International Conference on Robotics and Automation (ICRA), 2011* , 1–4 (May 2011).

[29] Barmpoutis, A., "Tensor body: Real-time reconstruction of the human body and avatar synthesis from rgb-d," *IEEE Transactions on Cybernetics* **43**(5), 1347–1356 (2013).